



## *User Guide*

---

# Affymetrix® GeneChip® Sequence Analysis Software Version 4.11

**For Research Use Only. Not for use in diagnostic procedures.**

## Trademarks

Affymetrix®, Axiom®, Command Console®, CytoScan®, DMET™, GeneAtlas®, GeneChip®, GeneChip-compatible™, GeneTitan®, Genotyping Console™, myDesign™, NetAffx®, OncoScan™, Powered by Affymetrix™, PrimeView™, Procarta®, and QuantiGene® are trademarks or registered trademarks of Affymetrix, Inc. All other trademarks are the property of their respective owners.

## Limited License

Subject to the Affymetrix terms and conditions that govern your use of Affymetrix products, Affymetrix grants you a non-exclusive, non-transferable, non-sublicensable license to use this Affymetrix product only in accordance with the manual and written instructions provided by Affymetrix. You understand and agree that, except as expressly set forth in the Affymetrix terms and conditions, no right or license to any patent or other intellectual property owned or licensable by Affymetrix is conveyed or implied by this Affymetrix product. In particular, no right or license is conveyed or implied to use this Affymetrix product in combination with a product not provided, licensed, or specifically recommended by Affymetrix for such use.

## Patents

Cartridge Array Software:

Products may be protected by one or more of the following patents: U.S. Patent Nos. 5,733,729; 5,795,716; 6,066,454; 6,090,555; 6,185,561; 6,188,783; 6,223,127; 6,228,593; 6,229,911; 6,308,170; 6,361,937; 6,420,108; 6,484,183; 6,505,125; 6,510,391; 6,532,462; 6,546,340; 6,567,540; 6,584,410; 6,611,767; 6,687,692; 6,826,296; 6,882,742; 6,957,149; 6,965,704; 6,996,475; 7,068,830; 7,130,458; 7,215,804; 7,424,368; 7,634,363; 7,822,555; 7,991,564; 7,992,098 and 8,190,373 and other U.S. and foreign patents.

## Copyright

©2005-2014 Affymetrix Inc. All rights reserved.



## TABLE OF CONTENTS

|                   |                                                           |           |
|-------------------|-----------------------------------------------------------|-----------|
| <b>Chapter 1</b>  | <b>Welcome . . . . .</b>                                  | <b>1</b>  |
|                   | Introduction . . . . .                                    | 1         |
|                   | Resequencing Array Types . . . . .                        | 2         |
|                   | About this Manual . . . . .                               | 2         |
| <b>Chapter 2</b>  | <b>Getting Started . . . . .</b>                          | <b>7</b>  |
|                   | The Workflow for Processing Resequencing Arrays . . . . . | 7         |
|                   | Requirements for Using GSEQ 4.1 . . . . .                 | 8         |
|                   | Starting GSEQ . . . . .                                   | 8         |
|                   | User Interface Components . . . . .                       | 11        |
| <b>Chapter 3</b>  | <b>Analyzing Cell Intensity Data . . . . .</b>            | <b>21</b> |
|                   | Introduction . . . . .                                    | 21        |
|                   | Performing Resequencing Analysis . . . . .                | 22        |
| <b>Chapter 4</b>  | <b>Resequencing Analysis Window . . . . .</b>             | <b>27</b> |
|                   | Overview . . . . .                                        | 28        |
|                   | Table View . . . . .                                      | 30        |
|                   | Sequence View . . . . .                                   | 40        |
|                   | SNP View . . . . .                                        | 48        |
|                   | Displaying Genomic and PCR Start/Stop Positions . . . . . | 48        |
| <b>Chapter 5</b>  | <b>Probe Intensity Window . . . . .</b>                   | <b>53</b> |
|                   | Overview . . . . .                                        | 54        |
|                   | Intensity Plot . . . . .                                  | 55        |
|                   | Probe Intensity Table . . . . .                           | 58        |
| <b>Chapter 6</b>  | <b>Resequencing Algorithm Reports . . . . .</b>           | <b>61</b> |
|                   | Resequencing Algorithm Report Contents . . . . .          | 61        |
| <b>Appendix A</b> | <b>Installing GSEQ . . . . .</b>                          | <b>67</b> |
|                   | Requirements . . . . .                                    | 67        |
|                   | Installation . . . . .                                    | 67        |

|                   |                                                         |           |
|-------------------|---------------------------------------------------------|-----------|
| <b>Appendix B</b> | <b>Resequencing Algorithm Version 1</b> . . . . .       | <b>71</b> |
|                   | Resequencing Algorithm Version 1 Description . . . . .  | 71        |
|                   | Resequencing Algorithm Version 1 Settings . . . . .     | 74        |
| <b>Appendix C</b> | <b>Resequencing Algorithm Version 2</b> . . . . .       | <b>77</b> |
|                   | Resequencing Algorithm Version 2 . . . . .              | 77        |
|                   | Resequencing Algorithm Version 2 Settings . . . . .     | 82        |
| <b>Appendix D</b> | <b>File Types</b> . . . . .                             | <b>85</b> |
|                   | Probe Information Files . . . . .                       | 85        |
|                   | Sample and Data Files . . . . .                         | 85        |
| <b>Appendix E</b> | <b>IUPAC Base Codes</b> . . . . .                       | <b>87</b> |
| <b>Appendix F</b> | <b>Working with Windowpanes &amp; Columns</b> . . . . . | <b>89</b> |
|                   | Resizing Windowpanes . . . . .                          | 89        |
|                   | Resizing or Hiding Table Columns . . . . .              | 90        |
| <b>Appendix G</b> | <b>Hot Key Descriptions</b> . . . . .                   | <b>91</b> |
| <b>Index</b>      |                                                         | <b>93</b> |

# WELCOME

Welcome to the Affymetrix GeneChip® Sequence Analysis Software (GSEQ) User's Guide. GSEQ is a software application that allows you to:

- Generate high quality sequence data from CustomSeq arrays
- Evaluate the data
- Export the data to other software applications for further study

## Introduction

Affymetrix GeneChip Sequence Analysis Software (GSEQ) 4.1 enables scientists to perform a variety of comparative sequencing tasks by taking advantage of the high-quality base calls and SNP calls produced. Because of the high accuracy of the base calls from each sample, researchers can save valuable time in data editing and complete projects more quickly.



**NOTE:** Directions for installing GSEQ are in [Appendix A, Installing GSEQ on page 67](#).

GSEQ Benefits include:

- Automated base calling with >99.9 percent accuracy
- Reporting of Call Rate by sample and by fragment
- Automatic assignment of homozygote positions with editing and edit tracking
- Detailed analysis of no calls through display of Force Call and no call criteria
- Alignment and display of sequences from multiple samples
- Display of user-defined genomic positions and PCR start/stop positions
- SNP Viewer with graphical display summarizing calls of samples at SNP sites
- Trace view allows visual display of probe intensities across sequence regions
- File Sets allows customized groupings of samples from multiple projects
- Flexible Export options: Export FASTA format as All Sequence or SNPs only

GSEQ is part of an integrated suite of software applications that allows you to:

- Perform different GeneChip probe array experiments
- Analyze the data from those experiments
- Organize the data
- Export the data to other software applications for further study

This chapter has the following sections:

- [Resequencing Array Types](#)
- [About this Manual](#)

## Resequencing Array Types

Resequencing Arrays provide an efficient and cost-effective method for resequencing large amounts of DNA. These arrays present the opportunity to sequence at least 300,000 bases of user-specified sequence on a single high-density array.

An Affymetrix® resequencing probe array consists of a number of probe cells. Each probe cell contains many copies of a unique 25-base oligonucleotide probe of defined sequence. Probe cells are grouped into probe sets to query a specific site in a known reference sequence.

Resequencing Algorithm Version 1 is used with 20 x 25  $\mu\text{m}$  arrays. It uses cell intensity data (with the .cel file extension) to make calls for every base position represented on the resequencing probe array. The algorithm uses the intensity data from multiple samples to improve its calling accuracy. The algorithm computes a quality score for each call as a metric of the call statistical accuracy and saves the calls and quality scores to an analysis results file. For more information about the algorithm, see [Appendix B, Resequencing Algorithm Version 1 on page 71](#).

Resequencing Algorithm Version 2 is used with 8  $\mu\text{m}$  arrays. It uses cell intensity data (with the .cel file extension) to make calls for every base position represented on the resequencing probe array. The algorithm uses the intensity data from multiple samples to improve its calling accuracy. The algorithm computes a quality score for each call as a metric of the call statistical accuracy and saves the calls and quality scores to an analysis results file. For more information about the algorithm, see [Appendix C, Resequencing Algorithm Version 2 on page 77](#).

The analysis results are displayed in the Resequencing Analysis window for evaluation and can be exported in formats that other software applications can use. For more information about the display of resequencing data, see [Chapter 4, Resequencing Analysis Window on page 27](#).

## About this Manual

This manual presents information about the GSEQ software in the following chapters and appendices:

- [Chapter 2, Getting Started](#): Describes the options for using GSEQ, how to start GSEQ, and the interface components.
- [Chapter 3, Analyzing Cell Intensity Data](#): Describes how to analyze cell intensity data for Resequencing, assays.
- [Chapter 4, Resequencing Analysis Window](#): Describes the display of analysis results for Resequencing assays.
- [Chapter 5, Probe Intensity Window](#): Describes how to view the probe intensity values for particular probe sets for Resequencing assays.
- [Chapter 6, Resequencing Algorithm Reports](#): Describes how to generate and view reports for the analyses performed in GSEQ.

The appendices include additional information and reference information:

- [Appendix A, Installing GSEQ](#): Provides installation instructions and information on initializing GSEQ.
- [Appendix B, Resequencing Algorithm Version 1](#): Describes the Resequencing algorithm, Version 1, and the user-changeable parameters.
- [Appendix C, Resequencing Algorithm Version 2](#): Describes the Resequencing algorithm, Version 2, and the user-changeable parameters.

The following appendices provide basic reference information for using GSEQ:

- [Appendix D, File Types](#)
- [Appendix E, IUPAC Base Codes](#)
- [Appendix F, Working with Windowpanes & Columns](#)

- [Appendix G, Hot Key Descriptions](#)

## Conventions Used in This Guide

This manual provides a detailed outline for all tasks associated with Affymetrix® GSEQ software. Various conventions are used throughout the manual to help illustrate the procedures described. Explanations of these conventions are provided below.

### Steps

Instructions for procedures are written in a numbered step format. Immediately following the step number is the action to be performed. Following the response, additional information pertaining to the step may be found and is presented in paragraph format. For example:

1. Click **Yes** to continue.

The Delete task proceeds.

In the lower right pane the status is displayed.

To view more information pertaining to the delete task, right-click **Delete** and select **View Task Log** from the shortcut menu.

### Font Styles

Bold fonts indicate names of commands, buttons, options or titles within a dialog box. When asked to enter specific information, such input opens in italics within the procedure being outlined. For example:

1. Click the **Find** toolbar button ; or

Select **Edit** → **Find** from the menu bar.

The Find dialog box opens.

2. Enter AFFX-BioB-5\_at in the **Find what** box, then click **Find Next** to view the first search result.
3. Continue to click **Find Next** to view each successive search result.

### Screen Captures

The steps outlining procedures are frequently supplemented with screen captures to further illustrate the instructions given.



**NOTE:** The screen captures depicted in this manual may not exactly match the windows displayed on your screen.

### Additional Comments

Throughout the manual, text and procedures are occasionally accompanied by special notes. These additional comments and their meanings are described below.



**TIP:** Information presented in Tips provide helpful advice or shortcuts for completing a task.



**NOTE:** The Note format presents supplemental information pertaining to the text or procedure being outlined.



**IMPORTANT:** The Important format presents important information that may affect the accuracy of your results.

---



**CAUTION:** Caution notes advise you that the consequence(s) of an action may be irreversible and/or result in lost data.

---



**WARNING:** Warnings alert you to situations where physical harm to person or damage to hardware is possible.

---

## On-line Documentation

The CD with the GSEQ software includes an electronic version of this guide. The electronic version is in Adobe® Acrobat® format (a \*.pdf file) and is readable with the Adobe® Acrobat Reader® software, available at no charge from Adobe at <http://www.adobe.com>.

The GSEQ software includes on-line help.

To access the on-line help:

- Click the **Help** button  in the main toolbar, or select **Help** → **Contents** from the main menu.

## Technical Support

Affymetrix provides technical support to all licensed users via phone or E-mail. To contact Affymetrix® Technical Support:

### **Affymetrix, Inc.**

3420 Central Expressway

Santa Clara, CA 95051 USA

E-mail: [support@affymetrix.com](mailto:support@affymetrix.com)

Tel: 1-888-362-2447 (1-888-DNA-CHIP)

Fax: 1-408-731-5441

### **Affymetrix UK Ltd.**

Voyager, Mercury Park

Wycombe Lane, Wooburn Green

High Wycombe HP10 0HH

United Kingdom

UK and Others Tel: +44 (0) 1628 552550

France Tel: 0800919505

Germany Tel: 01803001334

E-mail: [supporteurope@affymetrix.com](mailto:supporteurope@affymetrix.com)

Tel: +44 (0) 1628 552550

Fax: +44 (0) 1628 552585

**Affymetrix Japan, K. K.**

L O F U E ^j ^j ^pr`el ? iad) 4C

$\cdot \cdot / 1^5 E^j \wedge j \wedge \text{pr} \text{ `el) J fk}^d \cdot \text{hr}$

Tokyo . -2\*--.0

Japan

Tel: ( 5. \*%89\*310- \*1- / -

Fax: ( 5. \*%&\*310- \*1- / .

**www.affymetrix.com**



This chapter contains the following sections:

- [The Workflow for Processing Resequencing Arrays](#)
- [Requirements for Using GSEQ 4.1](#) on page 8
- [Starting GSEQ](#) on page 8
- [User Interface Components](#) on page 11

## The Workflow for Processing Resequencing Arrays

The processing of resequencing arrays follows the standard Affymetrix workflow. The steps from registering samples and probe arrays to generating cell intensity data files are handled by either Affymetrix GeneChip Command Console (AGCC) or GeneChip Operating System (GCOS), while the analysis of cell intensity data to generate probe analysis data is done in GSEQ 4.1 ([Figure 2.1](#)).

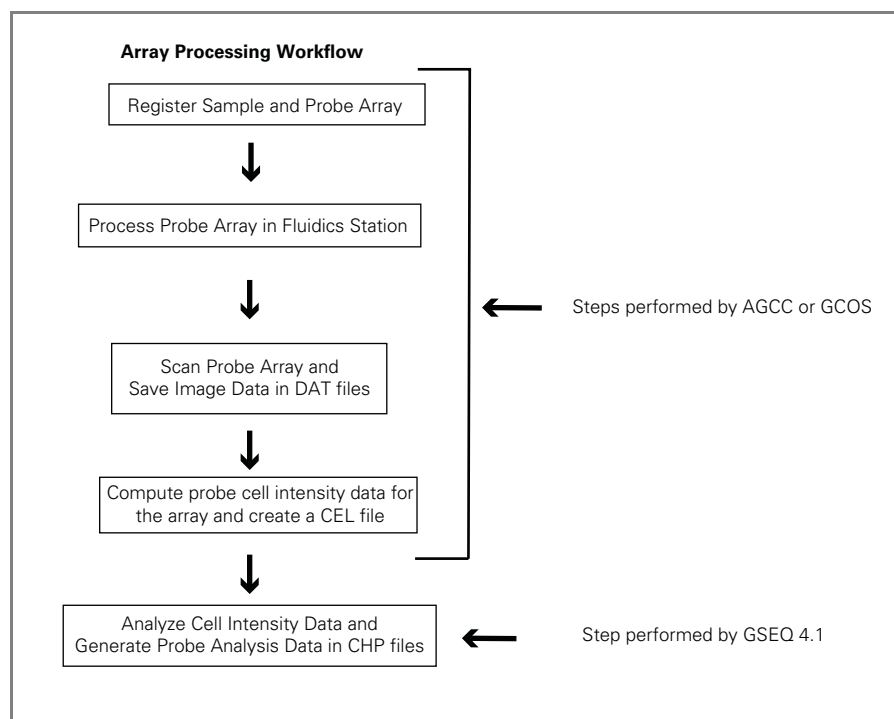


**Figure 2.1** Affymetrix GeneChip Workflow

The array processing workflow requires the tracking of sample and array information and of the array data. Information about the sample and array is handled in different ways in GCOS and AGCC. GCOS keeps the data in the Process Database, while AGCC stores the data in Sample (ARR) XML files.

A set of data files is produced for each array:

- Image (DAT) files: Contain pixel intensity values collected from an Affymetrix scanner, along with the gridding information used during feature extraction.
- Intensity (CEL) Data Files: Store the results of the intensity calculations on the pixel values of the DAT file.
- Probe Analysis (CHP) Files: Contain the probe analysis data for the array. CHP files are produced by the Analysis Application software and contain the actual data of interest.

GSEQ 4.1 takes the CEL file data produced by the array processing software and generates probe array data that can be exported to other software applications for further study.

There are differences between the data file formats used for AGCC and for GCOS. GSEQ 4.1 can work with CEL files produced by both GCOS and AGCC 1.1. The CHP files generated by GSEQ 4.1 are in AGCC format.

## Requirements for Using GSEQ 4.1

GSEQ can be run:

- On a computer with no other Affymetrix software
- On a computer with GCOS software
- On a computer with AGCC software

GSEQ 4.1 can work with CEL files in AGCC or GCOS format. The CHP files generated by GSEQ 4.1 are in AGCC format.

You must have the necessary library files installed to use GSEQ 4.1. If you do not have GCOS or AGCC on your computer, you can:

- Save GCOS files to a folder and specify that folder in the Library Path.
- Use the Library File Importer, installed with Data Exchange Console (DEC), to import library files to a folder and then specify that folder in the Library Path.

For information on specifying the Library Path, see [Changing the Paths for Data and Libraries on page 9](#).

For information on installing the DEC components, including the Library File Importer, and using the Library File Importer, see the *AGCC Installation Instructions*.

## Starting GSEQ

---

**!** **IMPORTANT:** You must have the necessary library files installed to use GSEQ 4.1.

---

**!** **IMPORTANT:** You will need to set up the paths for the data and library file folders, as described in [Changing the Paths for Data and Libraries on page 9](#).

---

**G** **NOTE:** Directions for installing GSEQ are in [Appendix A, Installing GSEQ on page 67](#).

---



**NOTE:** Affymetrix recommends that if you are using GCOS CEL files, you should use the Data Exchange Console provided with AGCC to copy the CEL files from the GCOS directory and to retain the sample and experiment attributes. If you choose not to copy your data to AGCC format, you may also use the Data Transfer Tool provided with GCOS.

#### To start GSEQ:

1. Click the Windows **Start** menu button and select **Programs** → **Affymetrix** → **GSEQ 4.1**.

At start up, the main window displays the data tree, shortcut bar, and status log (Figure 2.6).

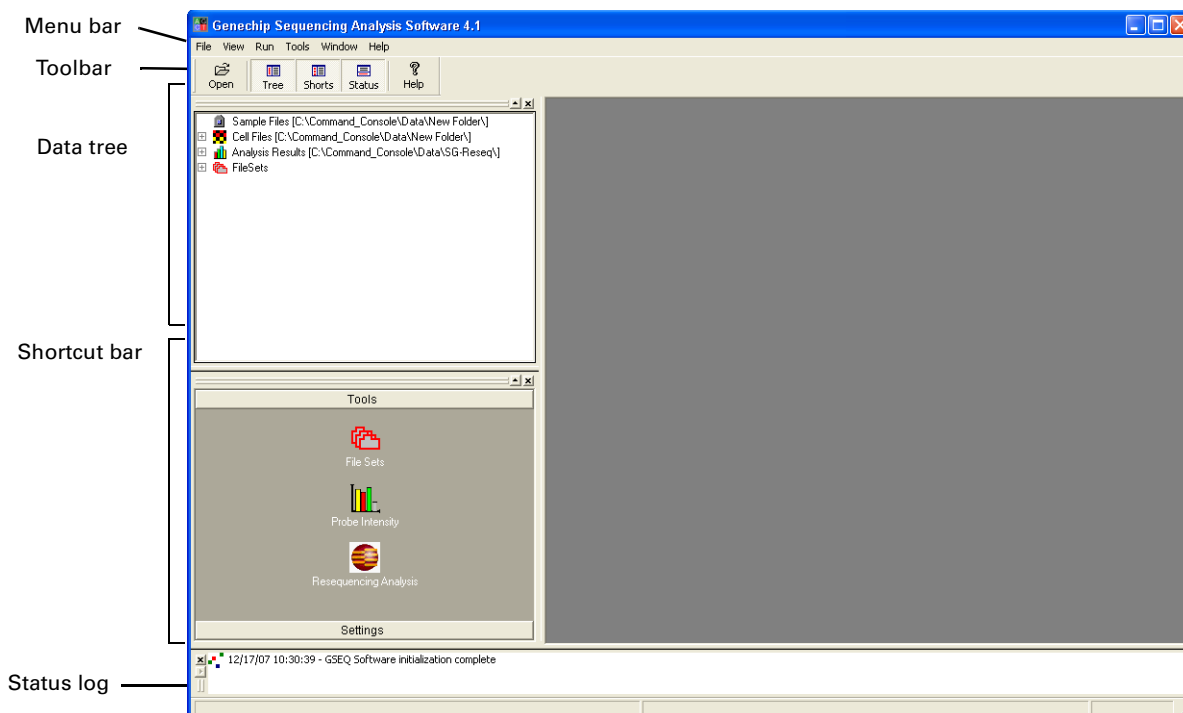


**Figure 2.2** GSEQ user interface at start up

Before using GSEQ, you need to set the data and library paths, as described below.

## Changing the Paths for Data and Libraries

#### To change the paths for Data and Libraries:

1. From the Tools menu, select **Defaults**; or  
In the Settings shortcut bar, select **Defaults**.  
The Set Data and Library Paths dialog box opens (Figure 2.3).

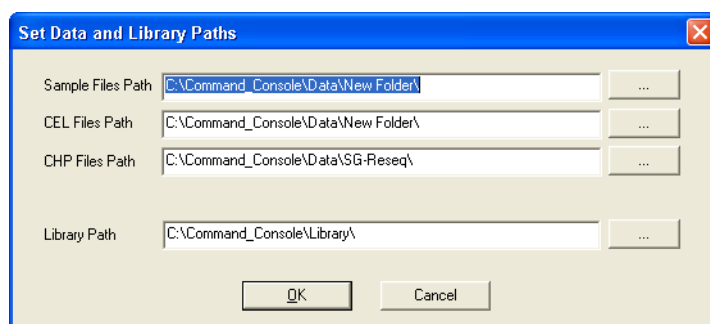


**Figure 2.3** Set Data and Library Paths dialog box

The dialog box displays boxes for the following:

- Sample Files path** Path to the folder where Sample files are located.
- CEL Files Path** Path to the folder where the CEL files are located.
- CHP Files Path** Path to the folder where the CHP files are located.
- Library Path** Path to the folder where the Library files for GSEQ are located.

2. Enter the path to the appropriate folder; or  
Click the Browse button to open the Data Folders dialog box.

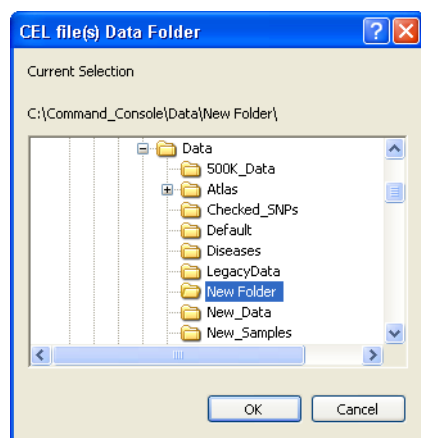


**Figure 2.4** CEL file(s) Data Folder dialog box

- A. You can then browse to the file location.
- B. Click **OK** in the Data Folders dialog box.
3. After selecting other folder paths, click **OK** in the Set Data and Library Paths dialog box.  
The paths to the folders are displayed in the data tree, along with any files.

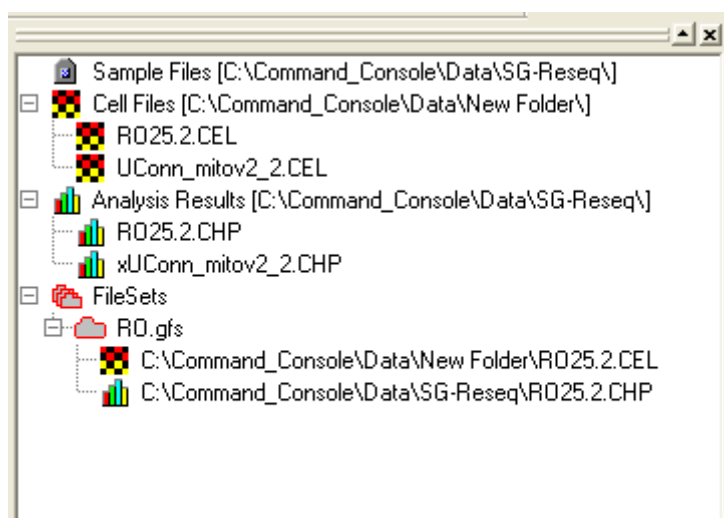


**Figure 2.5** Data tree with paths and files

See [Data Tree on page 13](#) for more information about using the data tree.

## User Interface Components

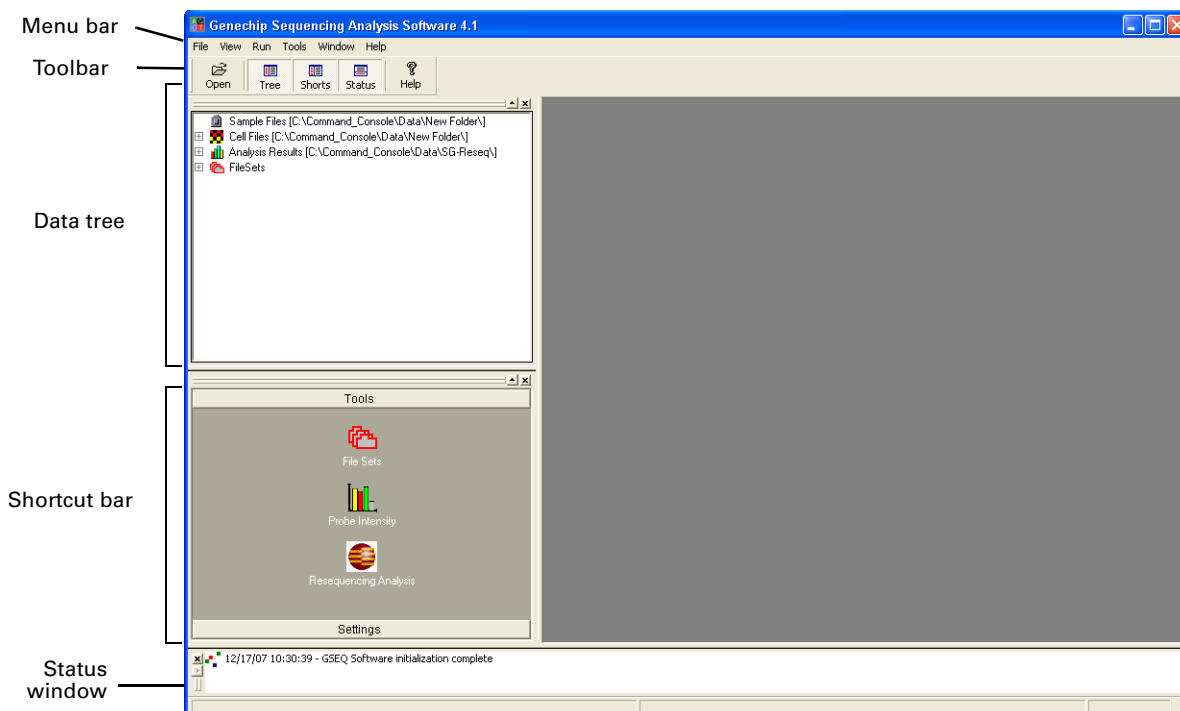


**Figure 2.6** GSEQ user interface at start up

The GSEQ user interface provides the following components:

- Data tree: Displays data available in GSEQ for the selected Assay type (see [Data Tree on page 13](#)).  
Use the data tree to:
  - View Project, Sample, Experiment, and FileSet Information.
  - Select cell intensity data for analysis.

- Select cell intensity data or analysis results for display.
- Main display area: Displays the following types of information:
  - Files selected for analysis (see [Performing Resequencing Analysis on page 22](#)).
  - Resequencing Analysis results (see [Chapter 4, Resequencing Analysis Window on page 27](#)).
  - Probe intensity data (see [Chapter 5, Probe Intensity Window on page 53](#)).
  - Generated reports (see [Chapter 6, Resequencing Algorithm Reports on page 61](#)).
- Shortcut bar, Tools View: Click the buttons to switch between open windows in the main display area, or to open Batch Analysis and Probe Intensity windows  
For more information, see [Tools Shortcut Bar on page 18](#).
- Shortcut bar, Settings: Click the buttons to open dialog boxes that allow you to change different default settings for the software.  
For more information, see [Settings Shortcut Bar on page 18](#).
- Status Window: Displays system messages about GSEQ activity.  
For more information, see [Status Window on page 18](#).
- Main Toolbar: Provides quick access to the commonly used functions of GSEQ software.

## Main Toolbar



**Figure 2.7** Main toolbar

You can display toolbars with text labels ([Figure 2.7](#)). To display the toolbar button labels, select **View** → **Toolbar** → **Text Labels** from the menu bar.

**Table 2.1** Main toolbar button functions

| Menu Bar Command       | Toolbar Button | Function                                                                                                          |
|------------------------|----------------|-------------------------------------------------------------------------------------------------------------------|
| File → Open            |                | Displays the Open dialog box so that a file (for example, analysis results or cell intensity data) may be opened. |
| Window → Data Tree     |                | Displays or hides the data tree.                                                                                  |
| Window → Shortcut Bar  |                | Displays or hides the shortcut bar.                                                                               |
| Window → Status Window |                | Displays or hides the status log.                                                                                 |
| Help → Contents        |                | Displays GSEQ help.                                                                                               |

## Data Tree

The data tree (Figure 2.8) displays the data in the selected folders organized by file type:

- Sample Files (files with .ARR extension)
- Intensity data (files with .CEL extension)
- Analysis Results data (files with .CHP extension)
- FileSets (groups of files with .gfs extension)

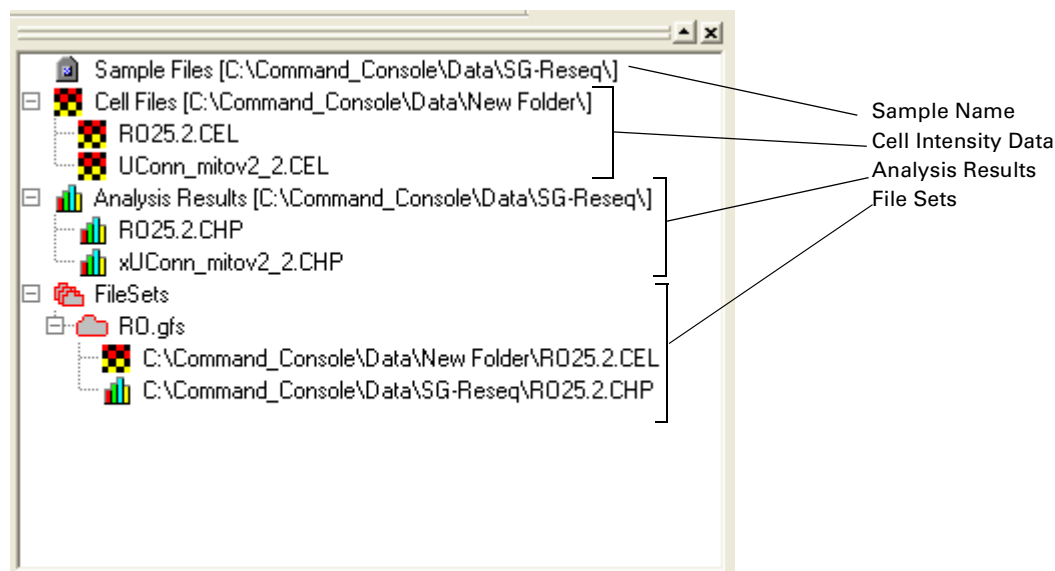


**Figure 2.8** GSEQ Data tree

In the data tree, you can perform the following tasks:

- View information about samples and data (see [Viewing File Information on page 14](#)).
- Open data for viewing (see [Opening Cell Intensity or Analysis Results Data on page 17](#)).
- Select cell intensity data for analysis (see [Performing Resequencing Analysis on page 22](#)).
- Rename FileSets (see page 14).

**To hide or display the data tree:**

- Click the **data tree** button  in the Main toolbar.

### Controlling the Display

You can expand and collapse tree components.

**To expand or collapse a tree component:**

- Click on the +/- button or right-click on the item and select the appropriate option from the shortcut menu.

**To hide a project altogether:**

- Right-click on the project and select **Hide Project(s)** from the shortcut menu.

## Viewing File Information

- In the data tree, right-click an item (sample file, cell intensity data, or analysis results) and select **Information** from the shortcut menu.

Information about the selected item(s) is displayed (Figure 2.9).

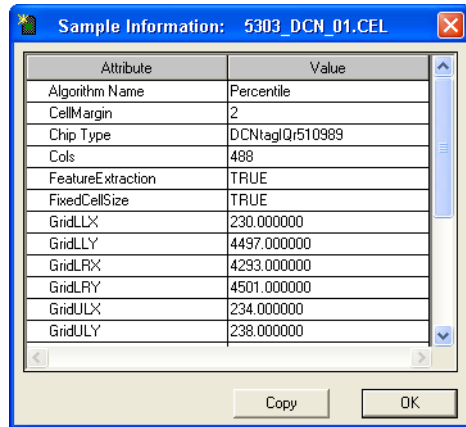


**Figure 2.9** CEL file information for Resequencing assay

## Using FileSets

FileSets allow you to create custom groups of GSEQ files (Figure 2.10). They allow you to manage large numbers of files more conveniently.

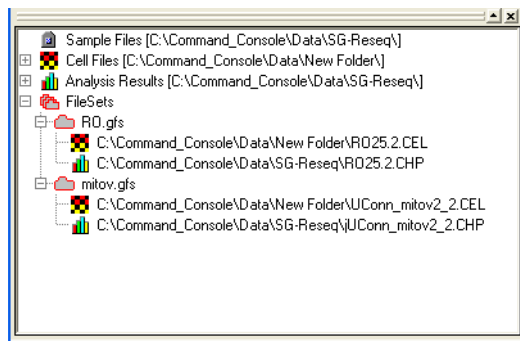


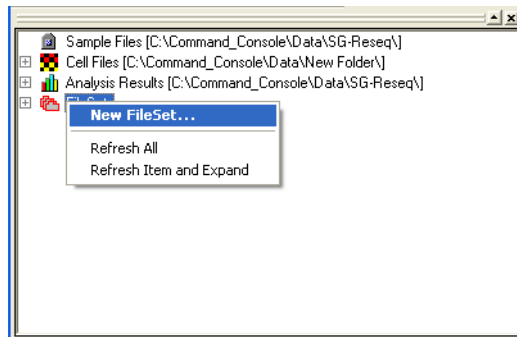
**Figure 2.10** FileSets in GSEQ Data Tree

You can create FileSets for any process where you need to select a file, for functions like:

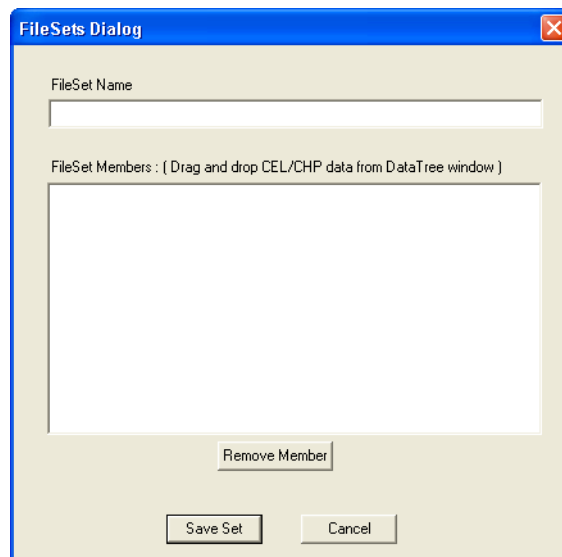
- Batch Analysis
- Displaying analysis results

**To group files into a FileSet:**

1. Right-click on the FileSets header in the Data Tree and select **New FileSets** from the shortcut menu (Figure 2.11); or

**Figure 2.11** FileSets header in Data Tree

From the **Tools** menu, select **File Sets**.  
The Fileset dialog box opens (Figure 2.12).

**Figure 2.12** FileSet dialog box

2. Select files for a set in the Data Tree, and drag them to the FileSet Members box (Figure 2.13).

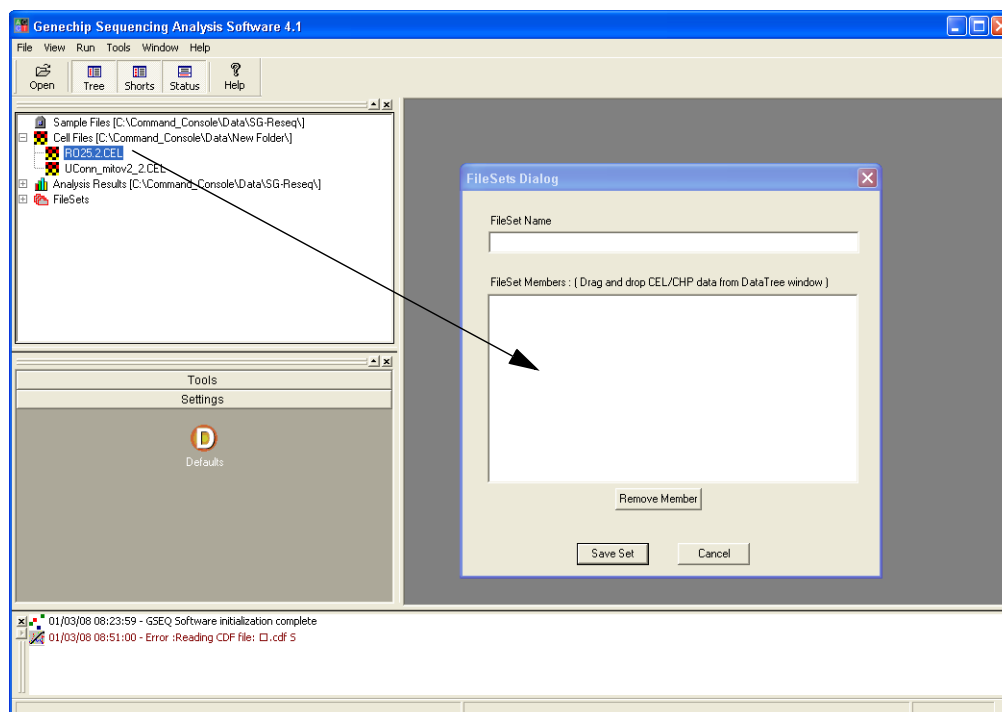


**Figure 2.13** Selecting files for a GSEQ FileSet

3. To remove a file from the Fileset:
    - A. Select the file in the FileSet Members list.
    - B. Click the Delete button.

The file is removed.
  4. Enter a name in the **FileSet Name** box.
  5. Click **Save Set** to create the FileSet.
- The FileSet appears in the Data Tree under the FileSets heading.

**To edit a fileset:**

1. Right-click on the FileSet in the data tree and select **Open FileSet** from the shortcut menu.
2. The FileSets dialog box opens and displays the selected fileset.
3. To add new files to the Fileset, drag them from the Data Tree.
4. To delete files from the fileset:
  - A. Selecting the file in the **FileSet Members** List.
  - B. Click the **Delete** button
5. Click **Save Set** to save the edited FileSet.

**To rename a fileset:**

1. Right-click on it in the Data Tree and select **Rename FileSet** from the shortcut menu.
2. The Rename Fileset dialog box appears (Figure 2.14).

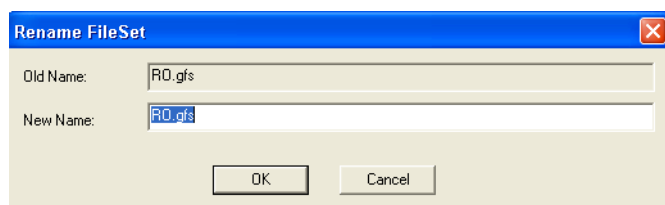


**Figure 2.14** Rename FileSet dialog box

3. Enter a new name in the **New Name** box and click **OK**.  
The fileset is renamed.

**To delete a fileset:**

1. Right-click on the fileset in the Data Tree and select **Delete** → **FileSet** from the shortcut menu.  
The Warning box appears (Figure 2.15).

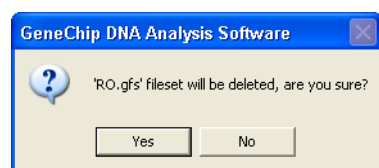


**Figure 2.15** Delete warning box

2. Click **Yes** to delete the fileset.

## Opening Cell Intensity or Analysis Results Data

You are able to display any data set by using the open command. Each data set has its own display, depending upon file type (cell intensity, analysis results, or report) and assay type.

1. In the data tree, do one of the following to open data:
  - Double-click the data.
  - Select the data, right-click the selection, and select **Open** in the shortcut menu.
  - Drag the selected data to the main display area.

The main display area displays the Resequencing analysis results in the Resequencing Analysis window. For more information, see [Chapter 4, Resequencing Analysis Window on page 27](#).

See [Resequencing Algorithm Report Contents on page 61](#) for information on opening an algorithm report.

## Shortcut Bar

The shortcut bar provides quick alternatives to menu bar commands. The shortcut buttons enable you to navigate between the open file types in the main display area. The shortcut bar has two sections:

- Tools (see below)
- Settings (see page 18)

**To hide or display the shortcut bar:**

- Click the **Shortcuts toolbar** button  in the Main toolbar.

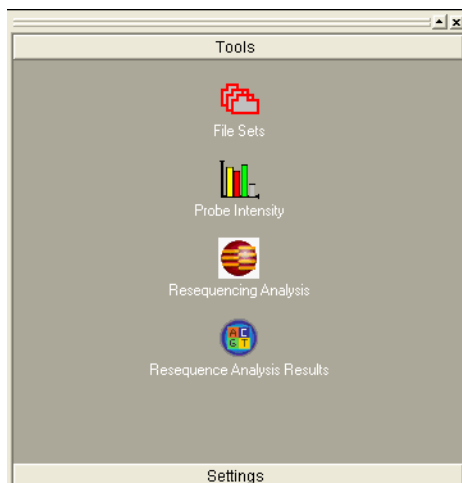
To toggle between the two shortcut bar sections:

- Click **Tools** or **Settings** in the shortcut bar.

## Tools Shortcut Bar

The buttons displayed in the Tools shortcut bar depend on the types of files that are open (Figure 2.16). If no files are open, the following buttons are available:

- File Sets
- Probe Intensity
- Resequencing Analysis



Click a button to change the view in the main display area:

Open the FileSet Dialog box

Open the Add Cell Files dialog box to select data to display in the Probe Intensity window or display an open Probe Intensity window.

Open the Resequencing Analysis window.

Display an open Resequencing Analysis Results window. (only displayed when the window is open)

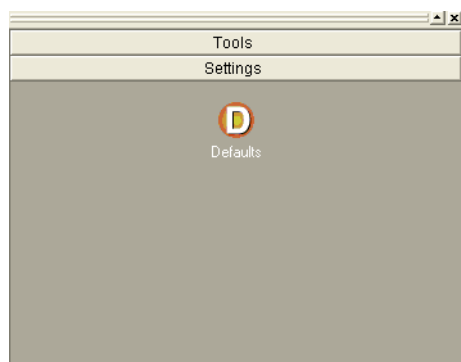
**Figure 2.16** Shortcut bar, Tools (GSEQ)

## Settings Shortcut Bar

The buttons in the Settings shortcut bar (Figure 2.17) provides easy access to the:

- Default dialog box, which allows you to set the data and library paths.

Click the button to perform the following function:



Open the Defaults dialog box.

**Figure 2.17** Shortcut bar, Settings (GSEQ)

## Status Window

The status window displays system status message (Figure 2.18).

- To clear the messages, right-click the status window and select **Clear Messages** from the shortcut menu.
- To hide (or display) the status window, click the **Status Log** button  in the Main toolbar.

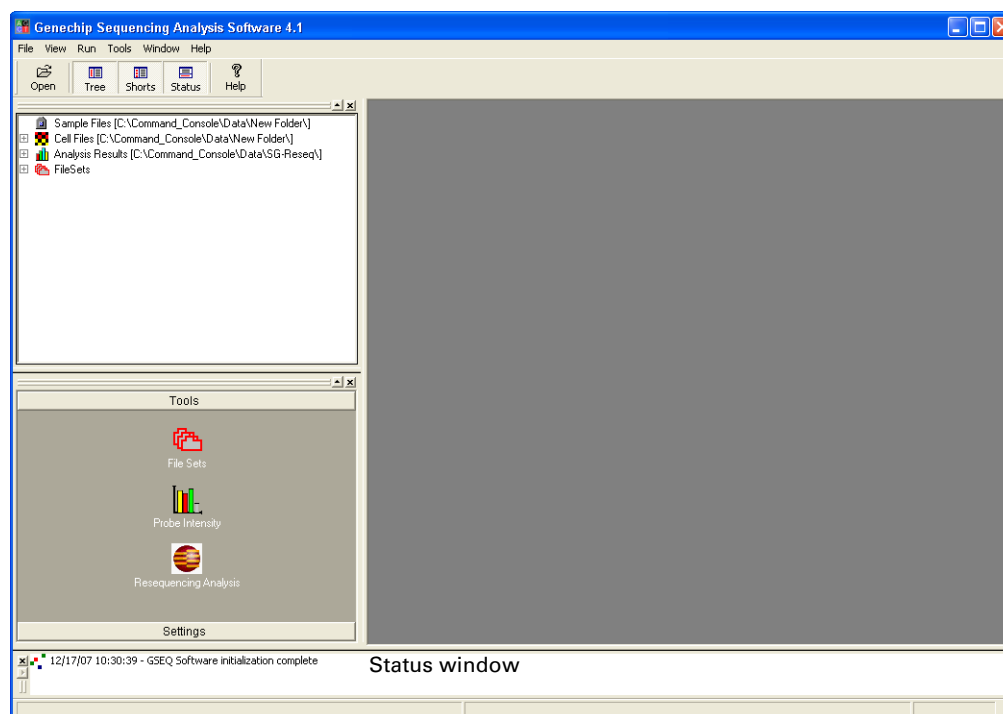


**Figure 2.18** GSEQ user interface, status window



The GSEQ Analysis function allows you to analyze cell intensity data from the Resequencing arrays (see page 21).

The analysis produces an analysis output file, with a .CHP file extension, for each sample. These output files can be displayed in the Resequencing Analysis windows.

For more information, see [Chapter 4, Resequencing Analysis Window on page 27](#).

When you perform Resequencing analyses, an algorithm report is automatically generated. Algorithm reports provide information about the algorithm performance and the samples analyzed.

For more information, see [Chapter 6, Resequencing Algorithm Reports on page 61](#).

## Introduction

You always follow the same basic steps:

1. Select the cell intensity data for analysis.
2. Assign names to the generated analysis results files and report files.  
If you don't assign new names to the analysis results and reports, previously created results and reports may be over-written.
3. Run the analysis.  
The proper algorithm is automatically selected for the chosen assay types.  
An algorithm report is generated for Resequencing analyses.



**NOTE:** GSEQ 4.1 can analyze CEL data files in either the GCOS or AGCC file format. The CHP files are created in AGCC format.

## Resequencing Analysis

For Resequencing arrays, GSEQ uses cell intensity data (in files with the .CEL extension) to make calls for every base position represented on the resequencing probe array. The algorithm uses the intensity data across multiple cell intensity data files to improve its calling accuracy and then computes a quality score for each call as a metric of the call statistical accuracy.

GSEQ has two resequencing algorithms:

- Resequencing Algorithm Version 1 is used with 20 x 25  $\mu\text{m}$  arrays.
- Resequencing Algorithm Version 2 is used with 8  $\mu\text{m}$  arrays.

GSEQ saves the calls and quality scores to an analysis results file (with the .CHP extension). The analysis results can be displayed in the Resequencing Analysis window (see [Chapter 4, Resequencing Analysis Window on page 27](#)). An algorithm report is also generated and can be displayed in the Report window (see [Chapter 6, Resequencing Algorithm Reports on page 61](#)).



**NOTE:** When performing a Resequencing analysis, for optimum algorithm performance, select 15 or more cell intensity data files from the same array type.



**IMPORTANT:** You may need to change the settings for model type, depending upon whether you are resequencing diploid or haploid samples. For more information, see:

- [Resequencing Algorithm Version 1 Settings on page 74](#)
- [Resequencing Algorithm Version 2 Settings on page 82](#)

For more information about Resequencing Algorithm Version 1 and the user-modifiable settings, see [Appendix B, Resequencing Algorithm Version 1 on page 71](#).

For more information about Resequencing Algorithm Version 2 and the user-modifiable settings, see [Appendix C, Resequencing Algorithm Version 2 on page 77](#).

## Performing Resequencing Analysis

The location of the CHP and Report files is set in the Set Data and Library Paths dialog box.

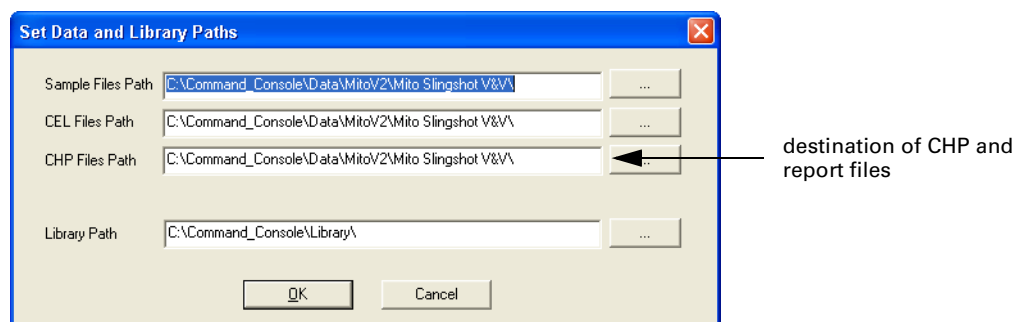


**Figure 3.1**

For more information, see [Changing the Paths for Data and Libraries on page 9](#).



**IMPORTANT:** You may need to change the settings for model type, depending upon whether you are resequencing diploid or haploid samples. For more information, see:

- [Resequencing Algorithm Version 1 Settings on page 74](#)
- [Resequencing Algorithm Version 2 Settings on page 82](#)

**To analyze intensity data using the Resequencing Analysis Window:**

1. Click the **Resequencing Analysis** shortcut button ; or

From the Run menu, select **Resequencing Analysis**

The Resequencing Analysis window opens ([Figure 3.2](#)).

This window enables you to:

- Select files for analysis
- Determine report name

**!** **IMPORTANT:** If you do not enter a new report name, an existing report may be overwritten without warning.

- Add prefix or suffix to CHP file names to prevent overwriting previously generated CHP files.
- Change the algorithm parameters for analysis (optional)
- Track progress of analysis.

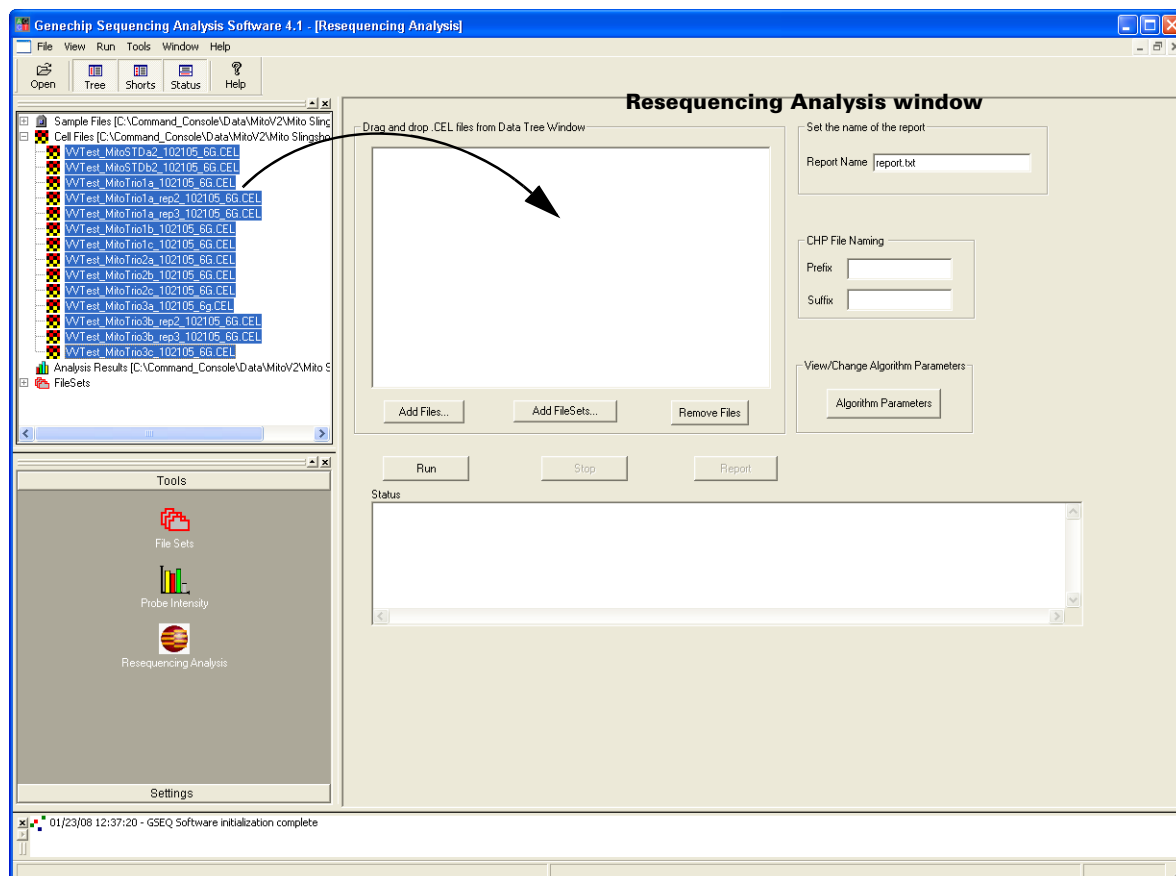


**Figure 3.2** Resequencing Analysis window in GSEQ

**G** **NOTE:** When performing a Resequencing analysis, for optimum algorithm performance, select 15 or more cell intensity data files from the same array type.

2. You can add the cell intensity data files to the Resequencing Analysis window using:

- The data tree
- The Add dialog box
- The File Sets feature

**Using the data tree:**

- A. In the data tree, select the cell intensity data files that you want to analyze (Figure 3.2).
- B. Drag the selection to the Resequencing Analysis window.

**Using the Add dialog box:**

- A. Click the **Add Files** button; or  
Select **Edit** → **Add Item** from the menu bar.

The Add Cell/Chip data items dialog box opens (Figure 3.3).

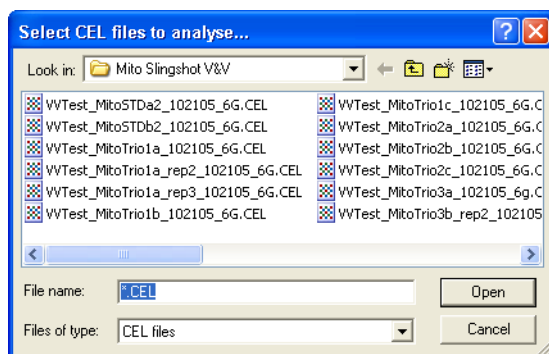


**Figure 3.3** Add Cell/Chip data items dialog box

- B.** Select the cell intensity data files that you want to analyze and click **Open**.  
The Resequencing Analysis window displays the selected data files (Figure 3.4).

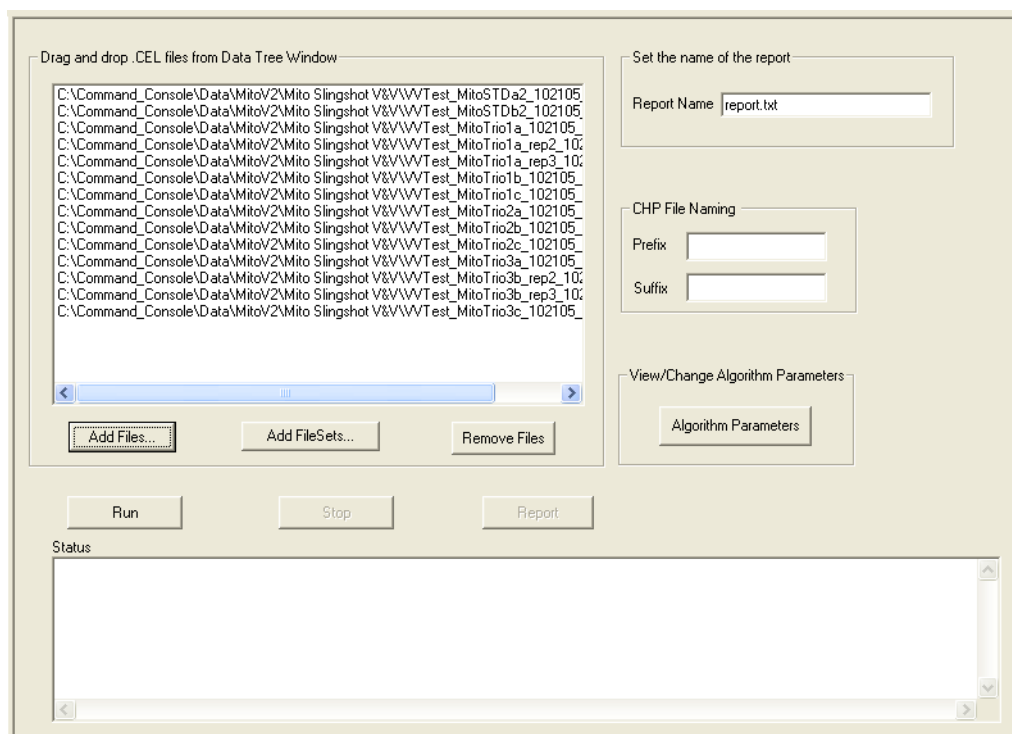


**Figure 3.4** Resequencing Analysis window with data from different types of resequencing arrays selected for analysis

### Select FileSets

- A.** Drag FileSets from the data tree to the Analysis window; or  
Click the **Add FileSets** button.  
The Select Sets Files dialog box opens (Figure 3.5).

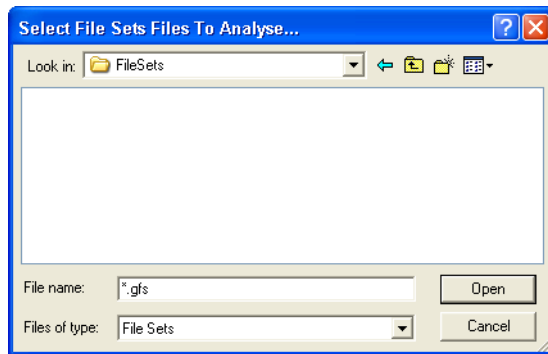


**Figure 3.5** Select Files Sets dialog box

- 1) Browse to the location with the file sets you wish to analyze.
- 2) Select the file sets and click **Open**.
3. To remove cell intensity data from the window:
  - Select the cell intensity data files and click the **Remove Files** button.
4. Enter a name for the algorithm report in the Report Name box (Figure 3.6).

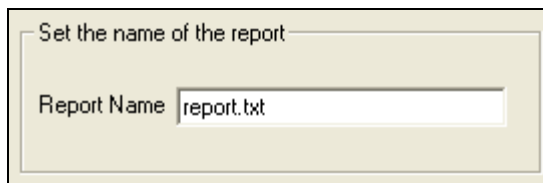


**Figure 3.6** Settings section



**IMPORTANT:** If you do not enter a new report name, an existing report may be overwritten without warning.

5. Enter a prefix or suffix for the resulting CHP file names (Figure 3.7).

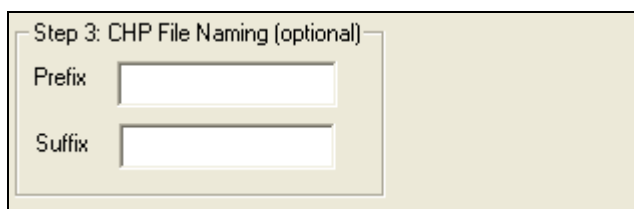


**Figure 3.7** Step 3: CHP File Naming (optional)

6. Click the Algorithm Parameters button to change the default parameters (optional).

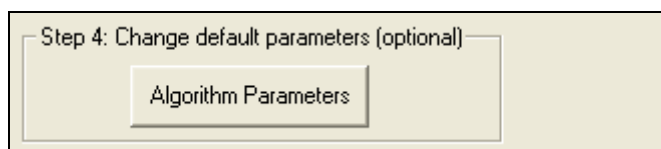


**Figure 3.8**

The appropriate Algorithm Parameters dialog box opens.

The parameters are different for:

- Resequencing Algorithm Version 1  
For more information, see [Resequencing Algorithm Version 1 Settings on page 74](#).
- Resequencing Algorithm Version 2  
For more information, see [Resequencing Algorithm Version 2 Settings on page 82](#).

**!** **IMPORTANT:** You may need to change the settings for model type, depending upon whether you are resequencing diploid or haploid samples. For more information, see:

- [Resequencing Algorithm Version 1 Settings on page 74](#)
- [Resequencing Algorithm Version 2 Settings on page 82](#)

7. Click the **Run** button.

If you do not have the analysis configuration file for your CustomSeq array, GSEQ will not be able to run the analysis and the following message will show up:

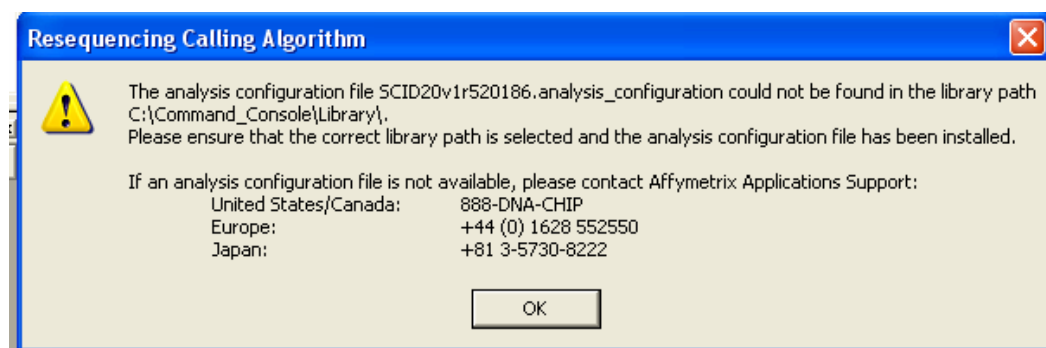


**Figure 3.9** Missing analysis configuration file notice

Follow the instructions in the notice before proceeding with the analysis.

If you are going to overwrite previously generated analysis results files, a warning box opens ([Figure 3.10](#)).

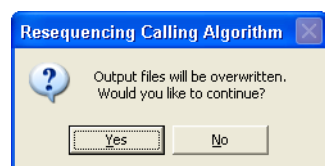


**Figure 3.10** Overwrite warning dialog box

- Click **Yes** to continue the analysis.

The appropriate analysis algorithm runs. When the analysis is completed, the appropriate algorithm report is displayed.

For more information about the various algorithm reports, see [Chapter 6, Resequencing Algorithm Reports on page 61](#).

**To stop an analysis:**

- Click the **Stop** button.



## Overview

When analyzing Resequencing assay data, GSEQ saves the base calls and associated quality scores for each sample to an analysis results file, with a .CHP file extension.

The Resequencing Analysis window displays the base calls in one of the following views:

- [Table View on page 30](#) (also with quality scores)
- [Sequence View on page 40](#)
- [SNP View on page 48](#)

## Displaying Resequencing Analysis Results

The Resequencing Analysis window opens automatically when you perform a resequencing analysis using the data tree or Analyze dialog box to select the cell intensity data files. You can open resequencing analysis results files for previously performed analyses using the:

- Data tree (see below)
- Open dialog box (see page 28)

### To open analysis results from the data tree:

1. Select the analysis results that you want to view.  
To select adjacent files, press and hold the **Shift** key while you click the first and last file in the selection. To select non-adjacent files, press and hold the **Ctrl** key while you click the files.
2. Right-click the selection and select **Open** in the shortcut menu.  
The Resequencing Analysis window displays the selected results.

### To open analysis results using the Open Dialog box:

1. Click the **Open** button ; or  
From the File menu, select **Open CEL/CHP**.  
The Open CEL/CHP dialog box opens ([Figure 4.2](#)).

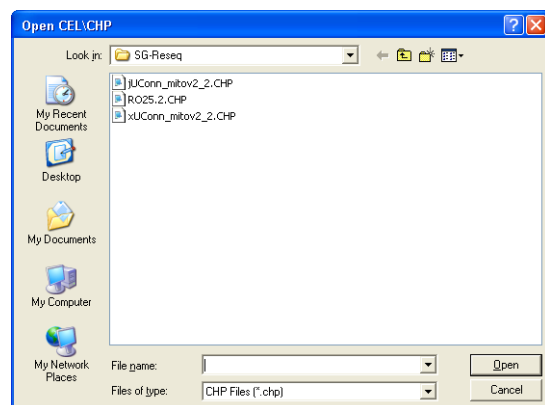


**Figure 4.2** Open dialog box

2. Select CHP Files (\*.chp) from the **Files of Type** drop-down list.
3. Select the analysis results data that you want to view and click **Open**.  
The Resequencing Analysis window displays the selected results.  
[right now we can only open one file at a time using this method.]

## Resequencing Analysis Window Toolbar

The toolbar (Figure 4.3) provides quick access to some of the commonly used functions of the Resequencing Analysis window.



**Figure 4.3** Resequencing Analysis window toolbar

You can display toolbars with text labels. To display the toolbar button labels, select **View → Toolbar → Text Labels** from the menu bar.

**Table 4.1** Resequencing Analysis window toolbar button functions

| Menu Bar Command            | Toolbar Button                                                                      | Function                                                                                                                                 |
|-----------------------------|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| Edit → Copy                 |    | Copies a highlighted selection in the table to the system clipboard (see page 36). Disabled in Sequence view.                            |
| Edit → Find                 |    | Opens the Find dialog box (see page 34).                                                                                                 |
| Export → FASTA              |   | Opens the FASTA Export dialog box (see page 35).                                                                                         |
| View → Option Settings      |  | Displays the Resequencing Analysis Window Options dialog box (see page 46). Disabled in Table view.                                      |
| View → Probe Intensities... |  | Opens the Add Cell Files dialog box, allowing you to drill down to view cell intensity data in the Probe Intensity window (see page 53). |

## Table View

In the Resequencing Analysis window, the table view ([Figure 4.4](#)) displays a list of the reference fragments in the probe arrays in the first column of the table.

| Fragment               | Chromosome | Chromosome Pos | Frag Pos | Tiling Pos | Ref   | Heterozygosity | Call | Quality   | Call | Quality   |
|------------------------|------------|----------------|----------|------------|-------|----------------|------|-----------|------|-----------|
| 3132 human_mtDNA_RC M1 | 2331       | 2329           | 3182     | t          | 0.000 | t              |      | 37.485237 | t    | 35.219636 |
| 3133 human_mtDNA_RC M1 | 2332       | 2330           | 3183     | c          | 0.000 | c              |      | 32.414024 | c    | 32.507988 |
| 3134 human_mtDNA_RC M1 | 2333       | 2331           | 3184     | c          | 0.000 | c              |      | 37.340988 | c    | 38.612762 |
| 3135 human_mtDNA_RC M1 | 2334       | 2332           | 3185     | g          | 0.000 | g              |      | 54.989784 | g    | 53.642654 |
| 3136 human_mtDNA_RC M1 | 2335       | 2333           | 3186     | c          | 0.000 | c              |      | 44.400497 | c    | 44.239357 |
| 3137 human_mtDNA_RC M1 | 2336       | 2334           | 3187     | a          | 0.000 | a              |      | 47.834511 | a    | 45.326645 |
| 3138 human_mtDNA_RC M1 | 2337       | 2335           | 3188     | t          | 0.000 | t              |      | 47.064476 | t    | 45.360405 |
| 3139 human_mtDNA_RC M1 | 2338       | 2336           | 3189     | a          | 0.000 | a              |      | 46.778587 | a    | 48.235184 |
| 3140 human_mtDNA_RC M1 | 2339       | 2337           | 3190     | a          | 0.000 | a              |      | 50.990318 | a    | 51.071632 |
| 3141 human_mtDNA_RC M1 | 2340       | 2338           | 3191     | g          | 0.000 | g              |      | 47.098381 | g    | 45.511665 |
| 3142 human_mtDNA_RC M1 | 2341       | 2339           | 3192     | c          | 0.000 | c              |      | 40.707321 | c    | 39.141022 |
| 3143 human_mtDNA_RC M1 | 2342       | 2340           | 3193     | c          | 0.000 | c              |      | 52.729568 | c    | 52.430855 |
| 3144 human_mtDNA_RC M1 | 2343       | 2341           | 3194     | t          | 0.000 | t              |      | 34.409515 | t    | 34.214867 |
| 3145 human_mtDNA_RC M1 | 2344       | 2342           | 3195     | g          | 0.000 | g              |      | 45.066811 | g    | 45.073090 |
| 3146 human_mtDNA_RC M1 | 2345       | 2343           | 3196     | c          | 0.000 | c              |      | 46.214058 | c    | 47.510666 |
| 3147 human_mtDNA_RC M1 | 2346       | 2344           | 3197     | g          | 0.000 | g              |      | 50.626228 | g    | 49.165497 |
| 3148 human_mtDNA_RC M1 | 2347       | 2345           | 3198     | t          | 0.000 | t              |      | 38.479698 | t    | 37.873970 |
| 3149 human_mtDNA_RC M1 | 2348       | 2346           | 3199     | c          | 0.000 | c              |      | 45.696571 | c    | 43.181107 |
| 3150 human_mtDNA_RC M1 | 2349       | 2347           | 3200     | a          | 0.000 | a              |      | 27.875404 | a    | 26.307610 |
| 3151 human_mtDNA_RC M1 | 2350       | 2348           | 3201     | g          | 0.000 | g              |      | 42.989433 | g    | 42.050514 |
| 3152 human_mtDNA_RC M1 | 2351       | 2349           | 3202     | a          | 0.000 | a              |      | 38.938461 | a    | 37.794212 |
| 3153 human_mtDNA_RC M1 | 2352       | 2350           | 3203     | t          | 0.000 | t              |      | 41.411644 | t    | 40.986612 |
| 3154 human_mtDNA_RC M1 | 2353       | 2351           | 3204     | t          | 0.000 | t              |      | 31.400047 | t    | 32.524429 |
| 3155 human_mtDNA_RC M1 | 2354       | 2352           | 3205     | a          | 0.000 | a              |      | 35.753807 | a    | 35.935867 |
| 3156 human_mtDNA_RC M1 | 2355       | 2353           | 3206     | a          | 0.000 | a              |      | 42.526985 | a    | 42.283768 |
| 3157 human_mtDNA_RC M1 | 2356       | 2354           | 3207     | a          | 0.000 | a              |      | 45.120308 | a    | 45.644493 |
| 3158 human_mtDNA_RC M1 | 2357       | 2355           | 3208     | a          | 0.000 | a              |      | 41.283539 | a    | 38.957520 |
| 3159 human_mtDNA_RC M1 | 2358       | 2356           | 3209     | c          | 0.000 | c              |      | 56.102570 | c    | 59.740944 |
| 3160 human_mtDNA_RC M1 | 2359       | 2357           | 3210     | a          | 0.000 | a              |      | 20.009552 | a    | 42.706393 |
| 3161 human_mtDNA_RC M1 | 2360       | 2358           | 3211     | c          | 0.000 | c              |      | 57.305901 | c    | 58.110947 |
| 3162 human_mtDNA_RC M1 | 2361       | 2359           | 3212     | t          | 0.000 | t              |      | 41.713028 | t    | 43.507072 |
| 3163 human_mtDNA_RC M1 | 2362       | 2360           | 3213     | a          | 0.000 | a              |      | 67.976707 | a    | 66.603639 |

**Figure 4.4** Resequencing Analysis window, Table view

For each fragment position, it displays:

|                            |                                                                                                                                                                                                                                                                                               |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Fragment</b>            | The reference fragment name described in the instruction file.                                                                                                                                                                                                                                |
| <b>Chromosome</b>          | The chromosome from which the reference fragment came (optional). See <a href="#">Displaying Genomic and PCR Start/Stop Positions</a> on page 48.                                                                                                                                             |
| <b>Chromosome Position</b> | The sequence position of the reference base in the chromosome (optional). See <a href="#">Displaying Genomic and PCR Start/Stop Positions</a> on page 48.                                                                                                                                     |
| <b>FragPos</b>             | The sequence position of the reference base in the reference fragment.                                                                                                                                                                                                                        |
| <b>Tiling Pos</b>          | The tile position number on the resequencing array of the probes for the reference base.                                                                                                                                                                                                      |
| <b>Ref</b>                 | The reference base at the fragment position.                                                                                                                                                                                                                                                  |
| <b>Heterozygosity</b>      | The relative frequency of mutations and reference base at a particular site in the samples. For example, heterozygosity=0.5 means a mutation and the reference base were called with equal frequency in the samples; heterozygosity=0 means no heterozygote calls were called in the samples. |

For each chp file, it also presents the following information:

|                    |                                                                                                                                                                                                       |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Call</b>        | The base call for the sample at the queried site.                                                                                                                                                     |
| <b>Forced call</b> | For a no call, the call that would have been made if the fail reason was ignored (Optional for Algorithm Version 2 only). See <a href="#">Displaying the Forced Call and Fail Reasons</a> on page 31. |

- Fail Reason** A code giving the reason a no call occurred (optional for Algorithm Version 2 only). See [Displaying the Forced Call and Fail Reasons on page 31](#).
- Quality** The total quality score computed for the base call. See [Appendix C, Resequencing Algorithm Version 2 on page 77](#).

You can:

- Display forced calls and fail reasons (Version 2 algorithm only) (see page 31)
- Display the genomic position (see [Displaying Genomic and PCR Start/Stop Positions on page 48](#))
- Sort the table (see below)
- Select columns to display or hide (see page 33)
- Search the table (see page 34)
- Export data in FASTA format (see page 35)
- Export data to a tab-formatted file (see page 35)
- [Copy data to the clipboard](#) for pasting into another application (see page 36)
- Edit the sequence information (see page 37)
- Drill down to view cell intensity data for a selected base in the [Probe Intensity window](#) (for more information, see [Displaying Probe Intensity Data on page 54](#))
- Print the data in the table (see page 39)

## Displaying the Forced Call and Fail Reasons

**! IMPORTANT: The forced call and fail reasons information is displayed only for arrays using the Version 2 Algorithm.**

A base call must pass several tests before it will be made and displayed in the Resequencing Analysis window as a good call. Otherwise it is displayed as a no call. In some cases a forced call can be made, even though the call does not meet reliability standards.

You can display a code describing why a call failed, and the forced call that would have been made, in the Table View ([Figure 4.5](#)).

|    | Fragment        | Chromosome | Chromosome Pos | Frag Pos | Tiling Pos | Ref | Heterozygosity | Call | Forced Call | Fail Reason |
|----|-----------------|------------|----------------|----------|------------|-----|----------------|------|-------------|-------------|
| 1  | mt440ref16159-1 | M          | 16200          | 40       | 43893      | t   | Undefined      | n    | t           | BR          |
| 2  | mt65ref16185-1E | M          | 16218          | 32       | 46963      | t   | Undefined      | n    | t           | BR          |
| 3  | mt54ref16161-1E | M          | 16225          | 63       | 45033      | t   | Undefined      | n    | t           | QS          |
| 4  | mt877ref161-212 | M          | 180            | 18       | 23933      | t   | Undefined      | n    | t           | BR          |
| 5  | mt223ref285-335 | M          | 313            | 27       | 31935      | t   | Undefined      | n    | t           | QS          |
| 6  | mt359ref16142-1 | M          | 16158          | 15       | 39046      | t   | Undefined      | n    | t           | BR          |
| 7  | mt61ref16165-1E | M          | 16199          | 33       | 46429      | t   | Undefined      | n    | t           | BR          |
| 8  | mt136ref16274-1 | M          | 16289          | 14       | 58821      | t   | Undefined      | n    | t           | BR          |
| 9  | mt9ref126-176   | M          | 155            | 28       | 21021      | t   | Undefined      | n    | t           | BR          |
| 10 | mt152ref16331-1 | M          | 16358          | 26       | 61216      | t   | Undefined      | n    | t           | QS          |

**Figure 4.5** Forced Call and Fail Reasons columns

### To display the forced calls:

- Select **View → Data Columns → Forced Call** from the main menu; or  
Right-click in the table and select **Data Columns → Forced Call** from the shortcut menu.

### To display the failure reasons:

- Select **View → Data Columns → Fail Reason** from the main menu; or  
Right-click in the table and select **Data Columns → Fail Reason** from the shortcut menu.  
The forced call and failure reason are displayed in the table for each chip.

The following codes are used to explain the call failure reasons:

|           |                                                                          |
|-----------|--------------------------------------------------------------------------|
| <b>NS</b> | No Signal Threshold                                                      |
| <b>WS</b> | Weak Signal Threshold                                                    |
| <b>SA</b> | Saturation Level                                                         |
| <b>QS</b> | Quality Score Threshold                                                  |
| <b>FP</b> | Foot Print--failed both Trace threshold and Sequence Profile             |
| <b>BR</b> | Base Reliability Threshold                                               |
| <b>DH</b> | The two strands both have Heterozygote calls on the two separate strands |

For more information about the resequencing algorithm and the failure, see [Resequencing Algorithm Version 2](#) on page 77.

## Sorting the Table

### Single Column Table Sort

You can sort the table by any data column in ascending or descending order.

- In the table, right-click the column header of interest and click **Sort Ascending** or **Sort Descending** in the shortcut menu.

### Multiple Column Table Sort

You can sequentially sort the table using up to four data columns.

1. Right-click the table and select **Sort** from the shortcut menu; or  
Select **Edit** → **Sort** from the menu bar.  
The Sort dialog box opens (Figure 4.6).

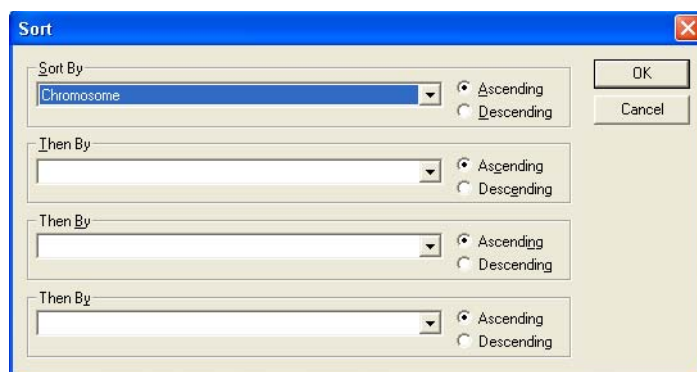


**Figure 4.6** Sort dialog box

2. Click the first (top) **Sort By** drop-down arrow, select a data column from the drop-down list, and choose the **Ascending** or **Descending** sort order (Figure 4.7).

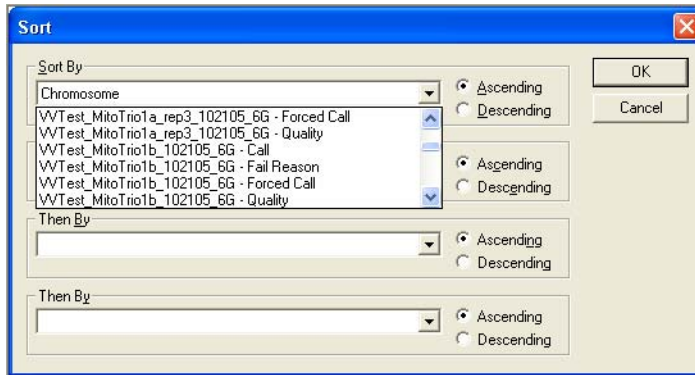


**Figure 4.7** Sort dialog box

3. To specify a second sort, click the drop-down arrow in the next **Then By** box, select a data column from the drop-down list, and choose the **Ascending** or **Descending** sort order.
4. To specify another sort, repeat step 3.
5. Click **OK**.

The table is arranged using the specified sort criteria.

## Hiding Columns

### To hide all **Call** (or **Quality**) columns:

- Right-click the table and select **Data Columns → Call** (or **Quality**) from the shortcut menu (Figure 4.8), or select **Data Columns → Call** (or **Quality**) from the menu bar.

### To display hidden **Call** (or **Quality**) columns:

- Right-click the table and select **Data → Columns Call** (or **Quality**) from the shortcut menu.

### You can also hide a user-selected column in the table:

- Right-click the header of the column that you want to hide and select **Hide Column** from the shortcut menu.

### To display hidden column(s):

- Right-click table and select **Show All Columns** from the shortcut menu.

| Copy Cols | Find              | PASTA    | Options    | PIW            |          |            |           |                |  |      | WTTest_Mt0Triol1a_rep3_102105_BG | WTTest_Mt0Triol1a_rep3_102105_BG | WTTest_Mt0Triol1a_rep3_102105_BG | WTTest_Mt0Triol1a_rep3_102105_BG |
|-----------|-------------------|----------|------------|----------------|----------|------------|-----------|----------------|--|------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
|           |                   | Fragment | Chromosome | Chromosome Pos | Frag Pos | Tiling Pos | Ref       | Heterozygosity |  | Call | Forced Call                      |                                  |                                  |                                  |
| 1         | mt440ref161591-M  |          | 16200      | 40             | 4389     | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 2         | mt65ref16185-M    |          | 16218      | 32             | 46363    | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 3         | mt54ref16161-M    |          | 16225      | 63             | 45033    | t          | Undefined | n              |  |      | t                                |                                  |                                  | QS                               |
| 4         | mt87ref16121-M    |          | 180        | 18             | 23933    | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 5         | mt223ref285-33-M  |          | 313        | 27             | 31935    | t          | Undefined | n              |  |      | t                                |                                  |                                  | QS                               |
| 6         | mt359ref16142-M   |          | 16158      | 15             | 39046    | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 7         | mt61ref16165-M    |          | 16199      | 33             | 46429    | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 8         | mt136ref16274-M   |          | 16289      | 14             | 58821    | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 9         | mt3ref126-176-M   |          | 155        | 28             | 21021    | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 10        | mt152ref16331-M   |          | 16358      | 26             | 61216    | t          | Undefined | n              |  |      | t                                |                                  |                                  | QS                               |
| 11        | mt266ref126-193-M |          | 156        | 29             | 20938    | t          | Undefined | n              |  |      | t                                |                                  |                                  | QS                               |
| 12        | mt10ref126-176-M  |          | 150        | 23             | 20701    | t          | Undefined | n              |  |      | t                                |                                  |                                  | QS                               |
| 13        | mt3ref126-176-M   |          | 153        | 26             | 21019    | c          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 14        | mt193ref201-25C-M |          | 218        | 16             | 26075    | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 15        | AFFX-TagID-EX     |          | 975        | 975            | 802      | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 16        | mt183ref165-224-M |          | 193        | 27             | 24734    | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 17        | mt777ref16274-M   |          | 16305      | 30             | 59029    | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 18        | mt187ref171-23C-M |          | 210        | 38             | 25587    | t          | Undefined | n              |  |      | t                                |                                  |                                  | QS                               |
| 19        | mt162ref158-22C-M |          | 205        | 46             | 21973    | t          | Undefined | n              |  |      | t                                |                                  |                                  | QS                               |
| 20        | mt194ref204-287-M |          | 224        | 19             | 26506    | t          | Undefined | n              |  |      | t                                |                                  |                                  | QS                               |
| 21        | mt830ref16331-M   |          | 16358      | 26             | 61298    | t          | Undefined | n              |  |      | t                                |                                  |                                  | QS                               |
| 22        | mt184ref171-22C-M |          | 209        | 37             | 25337    | t          | Undefined | n              |  |      | t                                |                                  |                                  | BR                               |
| 23        | mt193ref201-25C-M |          | 234        | 32             | 26091    | t          | Undefined | n              |  |      | t                                |                                  |                                  | QS                               |

**Figure 4.8** Resequence Analysis window, table view, with right-click menu

## Searching the Table

You can perform a text search on the cells in the table.

**To perform a text search:**

1. Click the **Find** button ; or  
Select **Edit** → **Find** from the menu bar.  
The Find dialog box opens (Figure 4.9).

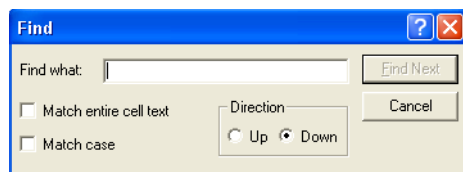


**Figure 4.9** Find dialog box

2. Enter the text string for the search (up to 256 alphanumeric characters).
3. Choose the Match Entire Cell Text option to find an exact match for the search string in a cell.

**NOTE:** If the option is selected, the text in a cell must completely agree with the text in the dialog box. If the option is not selected, the Find command finds all strings that match the text string for the search. For example, using the Find command to search for the text string 32323 would find chr25:32323 as well as other occurrences of 32323.

4. Choose the **Match** case or **Direction** option, and click **Find Next**.
5. To continue the search, click **Find Next** again.

## Exporting Resequencing Data

You can export the table data:

- As a tabular formatted text file (see page 35)
- As a FASTA format file (see page 35)

- By copying cell contents directly to the Clipboard (see page 36)

### Tabular Format

You can export all or a selected portion of the table to a tab-delimited text file.

#### To export a table to a tabular format:

1. Select the rows or columns you want to export. If you want to export all rows and columns, make no selection in the table.
2. Select **Export** → **Table** from the menu bar.  
The Export As dialog box opens (Figure 4.10).

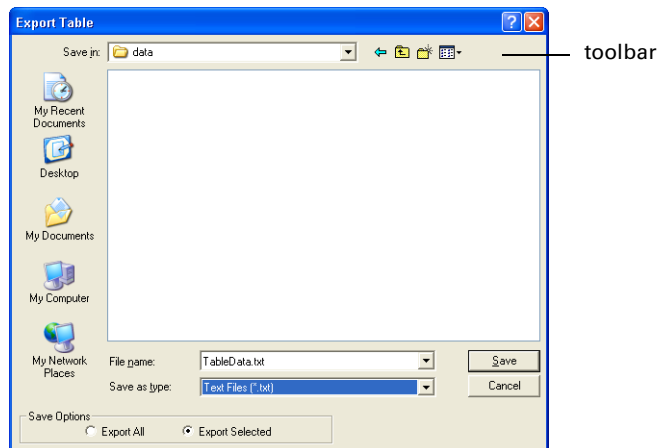


**Figure 4.10** Export Table dialog box

The dialog box toolbar provides standard options for selecting directories and displaying the list of files.

3. Select a directory for the file or use the default directory.
4. Enter a name for the \*.txt file.
5. Choose a Save Option:
  - **Export All:** exports the entire contents of the table.
  - **Export Selected:** exports the selected rows and columns.
6. Click **Save**.

The data is saved as a tabular format file.

### FASTA Format

#### To export data in FASTA format:

1. Click the **FASTA** button ; or  
Select **Export** → **FASTA** from the menu bar.  
The FASTA Export dialog box opens (Figure 4.11).

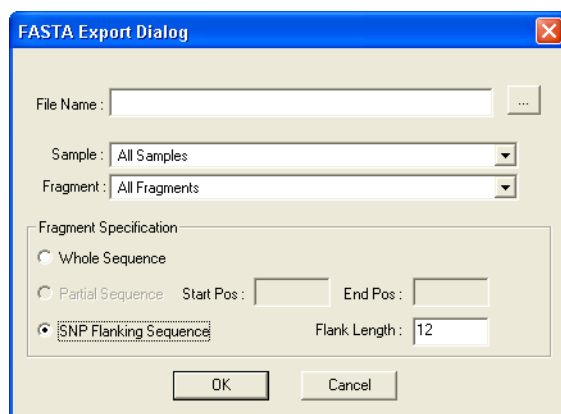


**Figure 4.11** FASTA Export dialog box

2. Enter a file name for the exported data.
3. Click the **Browse** button  to set the directory where the file will be saved.
4. Select an analysis results file (\*.chp) from the **Sample** drop-down list.
5. Select a fragment from the drop-down list. To export sequence information for all fragments, select **All** in the Fragment drop-down list.
6. To export the SNP calls with flanking sequences for the selected samples, select the **SNP Flanking Sequence** option and enter the flanking sequence length in the Flank Length edit box.
7. If you specified one fragment for export, choose the **Whole Sequence** or **Partial Sequence** option.
8. To export a partial or SNP sequence, enter the start and end positions.
9. Click **OK** to start the export.

### Copying the Table

All or part of the table can be copied to the system clipboard and pasted into other applications.

1. To select the entire table, click the upper left corner of the table ([Figure 4.12](#)).

|    | Fragment        | Chromosome | Chromosome Pos | Frag Pos | Tiling Pos | Ref | Heterozygosity | WTTest_MitoTrio1a_rep3_102105_6G | WTTest_MitoTrio1a_rep3_102105_6G | WTTest_MitoTrio1a_rep3_102105_6G |
|----|-----------------|------------|----------------|----------|------------|-----|----------------|----------------------------------|----------------------------------|----------------------------------|
|    |                 |            |                |          |            |     |                | Call                             | Forced Call                      | Call                             |
| 3  | mt54ref16161-16 | M          | 16225          | 63       | 45033      | t   | Undefined      | n                                | t                                | QS                               |
| 4  | mt877ref161-213 | M          | 180            | 18       | 23933      | t   | Undefined      | n                                | t                                | BR                               |
| 5  | mt223ref285-339 | M          | 313            | 27       | 31935      | t   | Undefined      | n                                | t                                | QS                               |
| 6  | mt359ref16142-1 | M          | 16158          | 15       | 39046      | t   | Undefined      | n                                | t                                | BR                               |
| 7  | mt61ref16165-16 | M          | 16199          | 33       | 46429      | t   | Undefined      | n                                | t                                | BR                               |
| 8  | mt136ref16274-1 | M          | 16289          | 14       | 58821      | t   | Undefined      | n                                | t                                | BR                               |
| 9  | mt9ref126-176   | M          | 155            | 28       | 21021      | t   | Undefined      | n                                | t                                | BR                               |
| 10 | mt152ref16331-1 | M          | 16358          | 26       | 61216      | t   | Undefined      | n                                | t                                | QS                               |
| 11 | mt266ref126-183 | M          | 156            | 29       | 20938      | t   | Undefined      | n                                | t                                | QS                               |
| 12 | mt10ref126-176  | M          | 150            | 23       | 20701      | t   | Undefined      | n                                | t                                | QS                               |
| 13 | mt9ref126-176   | M          | 153            | 26       | 21019      | c   | Undefined      | n                                | t                                | BR                               |
| 14 | mt193ref201-250 | M          | 218            | 16       | 26075      | t   | Undefined      | n                                | t                                | BR                               |
| 15 | AFFX-TagIQ-EX   |            | 975            | 975      | 802        | t   | Undefined      | n                                | t                                | BR                               |
| 16 | mt183ref165-224 | M          | 193            | 27       | 24734      | t   | Undefined      | n                                | t                                | BR                               |
| 17 | mt777ref16274-1 | M          | 16305          | 30       | 59029      | t   | Undefined      | n                                | t                                | BR                               |
| 18 | mt187ref171-239 | M          | 210            | 38       | 25587      | t   | Undefined      | n                                | t                                | QS                               |
| 19 | mt162ref158-223 | M          | 205            | 46       | 21973      | t   | Undefined      | n                                | t                                | QS                               |
| 20 | mt194ref204-287 | M          | 224            | 19       | 26506      | t   | Undefined      | n                                | t                                | QS                               |
| 21 | mt830ref16331-1 | M          | 16358          | 26       | 61298      | t   | Undefined      | n                                | t                                | QS                               |
| 22 | mt184ref171-222 | M          | 209            | 37       | 25337      | t   | Undefined      | n                                | t                                | BR                               |
| 23 | mt193ref201-250 | M          | 234            | 32       | 26091      | t   | Undefined      | n                                | t                                | QS                               |
| 24 | mt572ref16224-1 | M          | 16250          | 25       | 49652      | c   | Undefined      | n                                | t                                | BR                               |
| 25 | mt3ref119-176   | M          | 147            | 27       | 19889      | c   | Undefined      | n                                | t                                | QS                               |

Figure 4.12 Table view, all rows selected

## 2. To select part of the table, do one of the following:

- Use the click-and-drag method to select the rows.
- Click a row (or column) header to select the row (or column).

To select adjacent rows (or columns), press and hold the **Shift** key while you click the first and last row (or column) in the selection. To select non-adjacent rows (or columns), press and hold the **Ctrl** key while you click the rows (columns).

## 3. To copy the selection to the system clipboard, do one of the following:

- Click the **Copy Cells** button .
- Right-click the table and select **Copy Cells** from the shortcut menu.
- Select **Edit** → **Copy** from the menu bar.

The selected table cells are copied to the system clipboard.

## Editing Sequence Information

You can edit sequence information for a sample and save the edited information in a new analysis results file.

Editing sequence information requires two steps:

1. Edit the sequence calls for a sample (see below).
2. [Save the edited sequence calls](#) in a new analysis results file (see page 39).



**NOTE:** If you do not rename the edited analysis results file, your edits may be lost if you perform another analysis on the sample's cell intensity data file.



**NOTE:** You can restore an analysis results file to its unedited state by performing a resequencing analysis using the same algorithm parameters that were used in the original analysis.

**To edit a sequence call:**

1. Double-click in the cell for the fragment position row and the sample call column for the call you wish to edit (Figure 4.13).

Call Column

Fragment Position row

|    | Fragment            | Chromosome | Chromosome Pos | Frag Pos | Tiling Pos | Ref | Heterozygosity | Call | Forced Call |
|----|---------------------|------------|----------------|----------|------------|-----|----------------|------|-------------|
| 7  | mt61ref16165-16216  | M          | 16199          | 33       | 46429      | t   | Undefined      | n    | t           |
| 8  | mt136ref16274-16335 | M          | 16289          | 14       | 58821      | t   | Undefined      | n    | t           |
| 9  | mt9ref126-176       | M          | 155            | 28       | 21021      | t   | Undefined      | n    | t           |
| 10 | mt152ref16331-16386 | M          | 16358          | 26       | 61216      | t   | Undefined      | n    | t           |
| 11 | mt268ref126-183     | M          | 156            | 29       | 20938      | t   | Undefined      | n    | t           |
| 12 | mt10ref126-176      | M          | 150            | 23       | 20701      | t   | Undefined      | n    | t           |
| 13 | mt9ref126-176       | M          | 153            | 26       | 21019      | c   | Undefined      | n    | t           |
| 14 | mt193ref201-250     | M          | 218            | 16       | 26075      | t   | Undefined      | n    | t           |
| 15 | AFFX-TagIQ-EX       |            | 975            | 975      | 802        | t   | Undefined      | n    | t           |

**Figure 4.13** Selecting a call for editing

2. Enter the new sequence call using a standard IUPAC code.
3. Press **Enter**.

If you selected a standard IUPAC code, the change is entered in the file.

If you do not use a letter in the IUPAC code, a warning box opens (Figure 4.14).

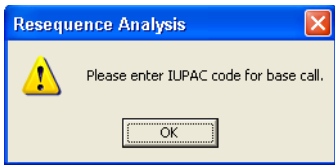


**Figure 4.14** IUPAC Warning box

- Click **OK** and enter a standard IUPAC code.
4. Select another call for editing.
  5. When you have finished editing the calls, save an analysis results file with the edits (see below).

**To restore an edited call to the original call:**

- Right-click in an edited cell and select **Restore to original calls for selected cells** from the short-cut menu.

The menu displays the original unedited call (Figure 4.15).

|       |   |           |   |   |
|-------|---|-----------|---|---|
| 20701 | t | Undefined | T | t |
| 21019 | c | Undefined | T | t |
| 26075 | t | Undefined | n |   |
| 802   | t | Undefined | n |   |
| 24734 | t | Undefined | n |   |
| 59029 | t | Undefined | n |   |
| 25587 | t | Undefined | n |   |
| 21973 | t | Undefined | n |   |
| 26506 | t | Undefined | n |   |
| 61298 | t | Undefined | n | t |
| 25337 | t | Undefined | n | t |

**Figure 4.15** Restoring an edited call

The original call is restored.

**To save the edits in an analysis results file:**

1. From the menu bar, select **Export** → **Save Calls**.  
The Save Chip dialog box opens (Figure 4.16).

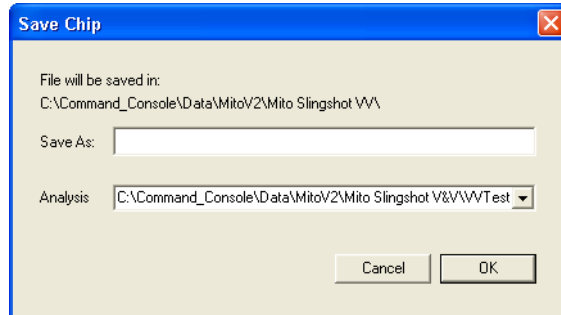


**Figure 4.16** Save Chip dialog box

2. Select the sample with the edits you wish to save from the Analysis drop-down menu.
3. Enter a name for the edited file in the Save As box.



**NOTE:** Using a unique file name for your edited data will protect it from being overwritten by subsequent resequencing analyses.

4. Click **OK** to save the edited data.  
The edited data is saved in an analysis results file.

You can view, export, and run reports using the edited sequence information, just as you can with an unedited analysis results file.



**TIP:** You may need to refresh the data tree to see the new file with the edited information. Right-click anywhere in the data tree and select **Refresh All** in the shortcut menu.

## Printing the Table Contents

**To print the table contents:**

1. Select the table cells you want to print.
2. Select **File** → **Print** from the menu bar.  
The Print dialog box opens (Figure 4.17).

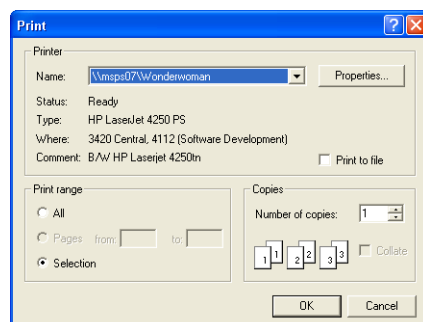


**Figure 4.17** Print dialog box for Resequencing Analysis window

The Print dialog box includes the standard Windows print options (select printer, change properties, number of copies, etc., depending upon your system and printer).

3. Select the print range radio button for the range you wish to print:
  - **All** prints the entire contents of the table.
  - **Selection** prints out only the contents selected in the table.



**NOTE: Selecting All will result in printing out many pages of table content.**

4. Click **OK**.

The table contents are printed.



**TIP: You may wish to experiment with the page layout and scaling properties of your printer (if available) to improve the printout of the table.**

## Sequence View

In the Resequencing Analysis window, the sequence view displays the sequence of the selected reference fragment and the base calls for selected Resequencing analysis results ([Figure 4.18](#)).

### To display the sequence view:

- Click the Sequence tab at the bottom of the Resequencing Analysis window.

The sequence view has four areas or windowpanes that can be resized ([Figure 4.18](#)). [Table 4.2](#) explains the display in each windowpane.

For more information about resizing windowpanes, see [Working with Windowpanes & Columns on page 89](#).

### To select the reference fragment in the sequence view:

- Click a radio button next to the fragment name (left windowpane).

**Table 4.2** Resequencing Analysis window, Sequence view windowpanes

| Windowpane Name              | Displays...                                                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Fragment pane<br>(left pane) | The names of the reference fragments represented on the resequencing probe array used for the selected analysis results. Choose a fragment in this windowpane for display in the view, map, and base windowpanes.                                                                                                                                                                                      |
| Scroll pane<br>(top pane)    | A representation of the selected reference fragment and the sequence viewer that enables you to specify all or part of the fragment for viewing in the map and base windowpanes. Use the scroll box in this pane to scroll through or zoom in or out on the map pane.                                                                                                                                  |
| Map pane<br>(middle pane)    | Color-coded sequence maps of the base calls for the samples. The scroll box in the scroll pane controls the region displayed in the map pane. Different colors highlight the reference calls, no calls, and SNP calls for easy identification. The view finder in the map pane shows the portion of the sequence map displayed in the base pane. Move the view finder to scroll through the base pane. |
| Base pane<br>(bottom pane)   | The base calls for the samples (chip files) for the selected fragment. The base pane displays the sequence inside the view finder.                                                                                                                                                                                                                                                                     |

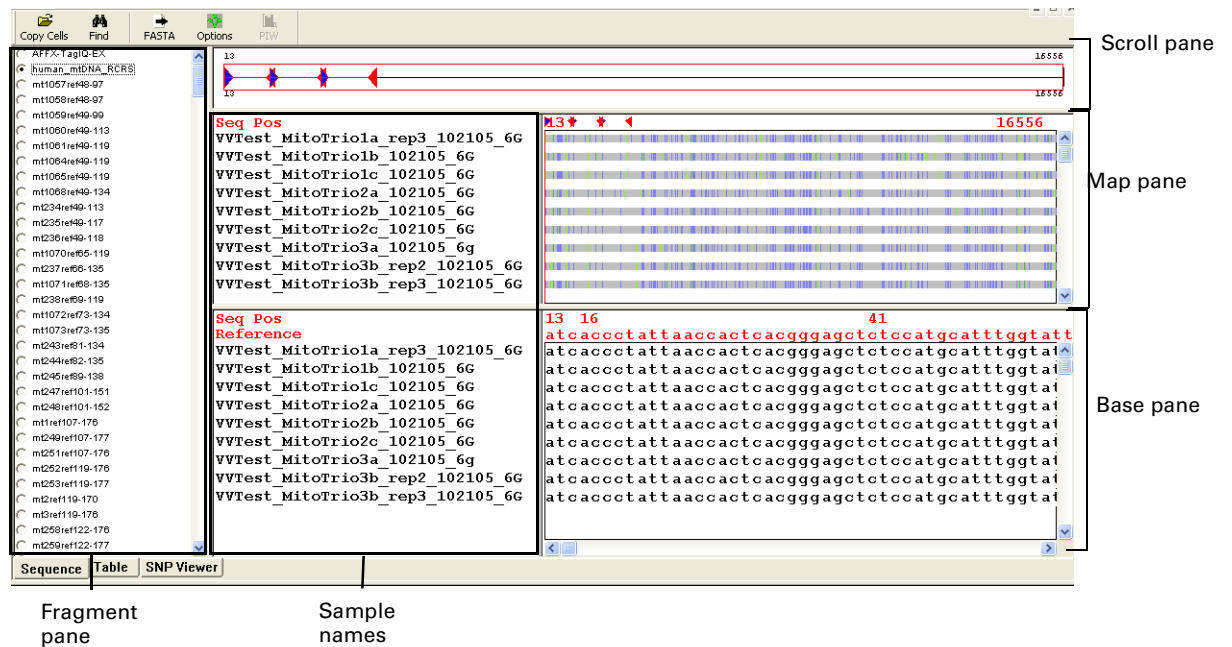


**Figure 4.18** Resequencing Analysis windowpanes, sequence view, with genomic positions and PCR start/stop positions displayed

You can:

- Navigate the Sequence View windowpanes (see below)
- Change display options in Sequence View (see page 46)
- Print the sequence map (see page 46)
- Export the sequence information in FASTA format (see page 35)

## Navigating the Sequence View Windowpanes

The scroll box determines the view in the map and base panes ([Figure 4.19](#)). Adjust the width of the scroll box and its position on the reference fragment to select the part of the reference sequence that you want to view. Decrease the width of the scroll box to zoom in on (magnify) the sequence map.



**Figure 4.19** Sequence view windowpanes

1. To adjust the width of the scroll box:
  - A. Place the mouse pointer over the left or right edge of the scroll box so that the pointer changes to a .
  - B. Click and hold the mouse button down and drag the edge to change the width of the scroll box. The map and base panes display the sequence inside the scroll box.

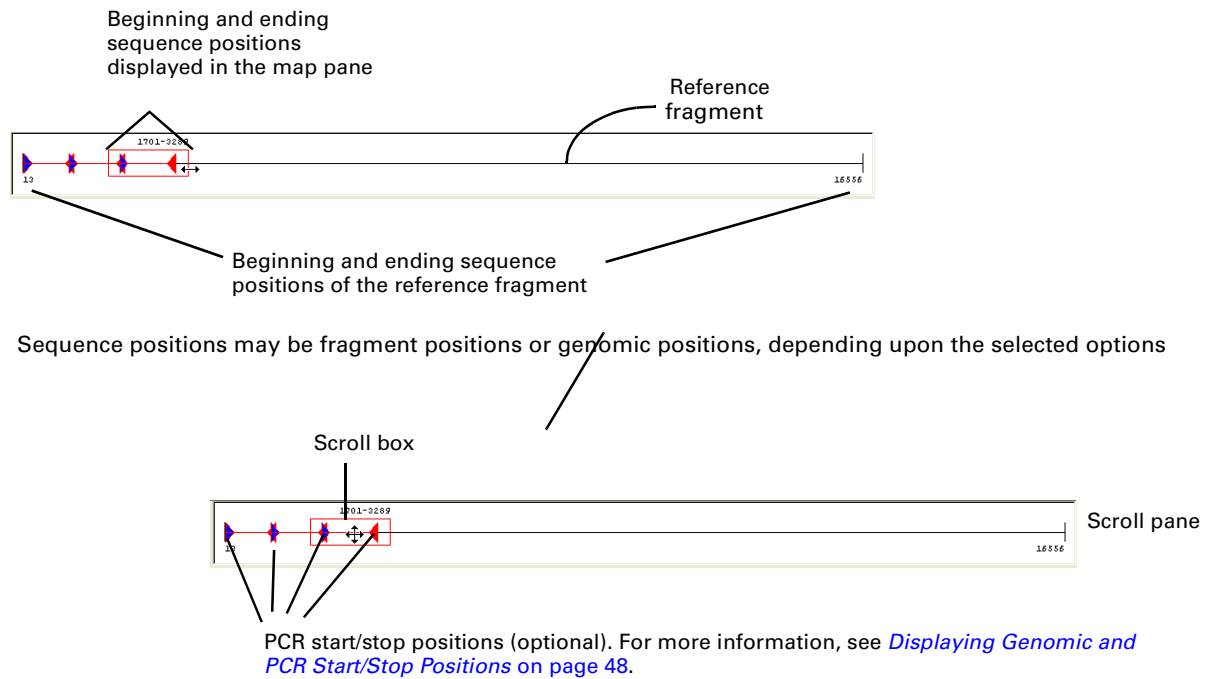


**Figure 4.20** Scroll pane, adjust length (top) or position (bottom) of the scroll box along the reference fragment to specify the display in the map pane

2. To adjust the position of the scroll box on the reference fragment:
  - A. Place the mouse pointer over the scroll box so that the pointer changes to a  (Figure 4.20).
  - B. Use the click-and-drag method to move the scroll box to the location in the reference fragment that you want to display in the map and base panes.

The map and base windowpanes display the sequence inside the scroll box (Figure 4.21).
3. To scroll through the base pane, move the view finder in the map pane or use the base pane scroll bars.

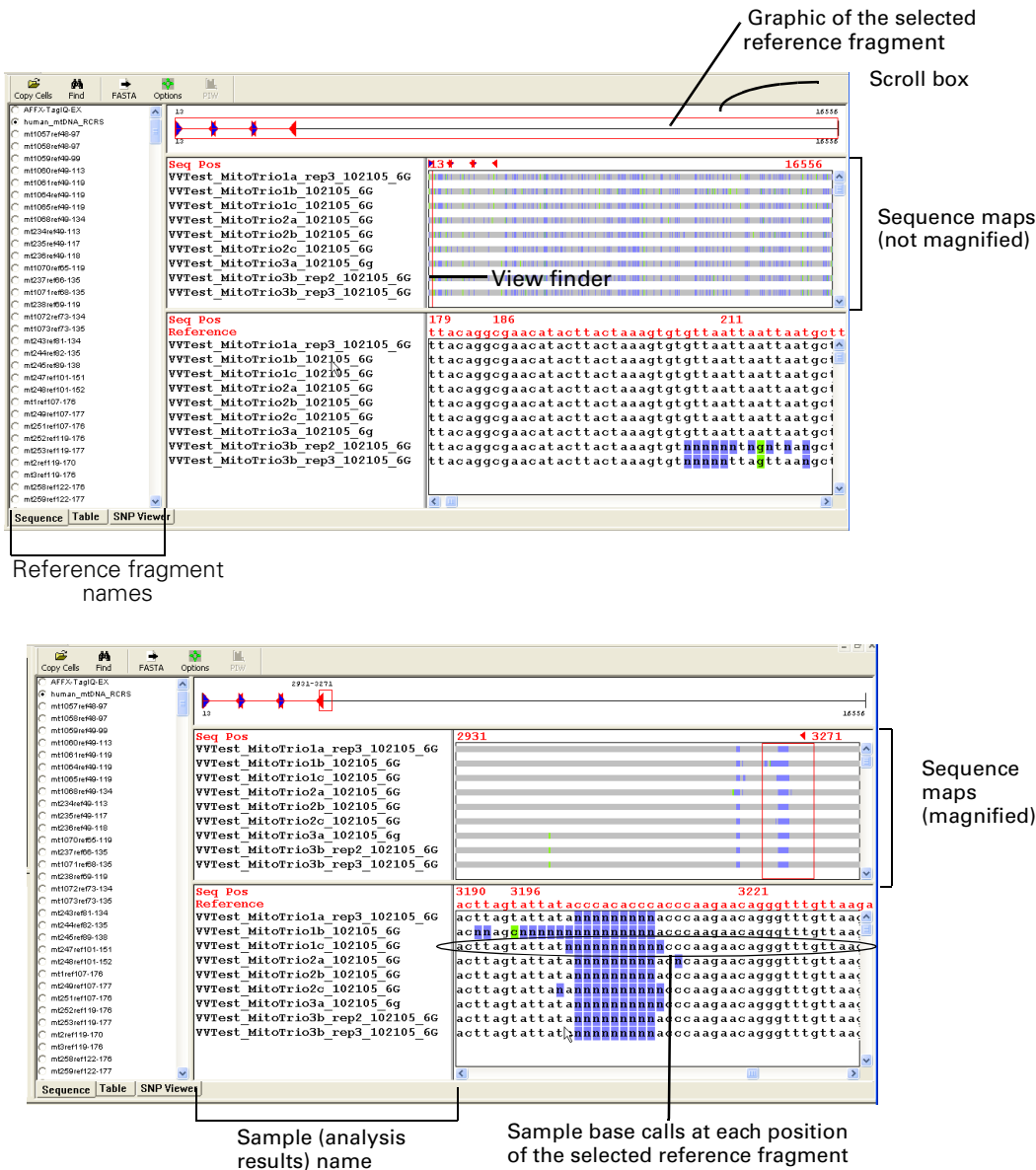


Figure 4.21 Sequence view, map pane sequence maps: not magnified (top) and magnified (bottom)

### Sequence Maps

The Sequence map pane displays color-coded sequence maps of the sample base calls. It shows the portion of the sequence inside the scroll box. The view finder shows the map area displayed in the base pane.

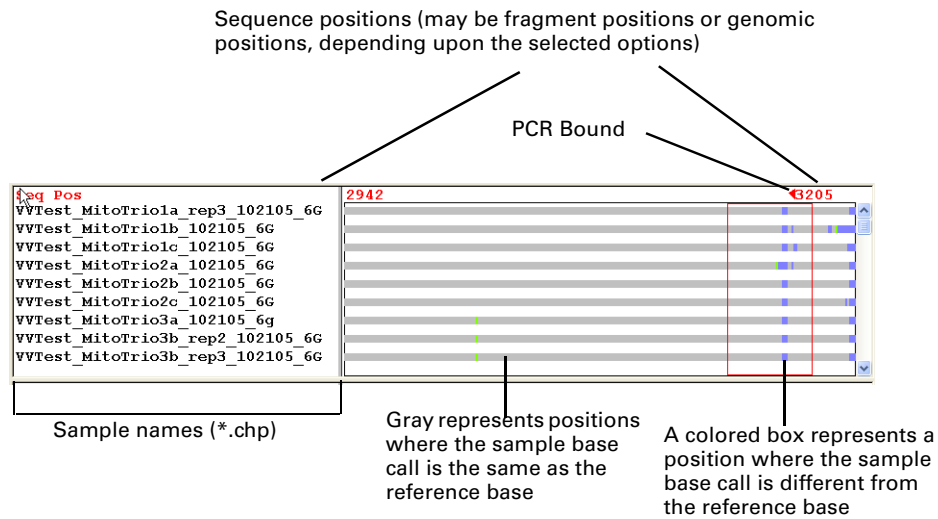


**Figure 4.22** Map windowpane

The Sequence map is color-coded for easy identification of the positions. The default position colors are:

- Gray** The sample base call is the same as the reference base.
- Blue** A no call.
- Green** A homozygous difference call.
- Orange** A heterozygous difference call.

**To scroll through the Sequence map pane, use one of the following methods:**

- Press the left or right arrow key (moves the view one base at a time left or right).
- Press the Page Up or Page Down key (moves the view finder to the next group of sequence maps above or below).

### Sample Base Calls

The base windowpane displays the selected reference sequence and the sample base calls (Figure 4.23). It shows sequence information for the portion of the fragment inside the scroll box. (See [IUPAC Base Codes](#) on page 87)

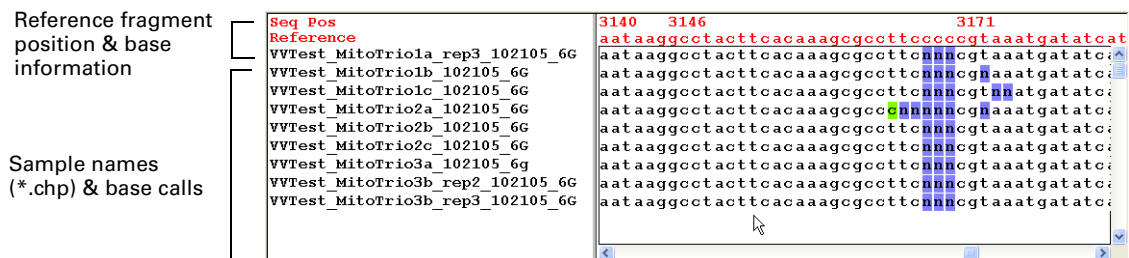


**Figure 4.23** Base pane

**To scroll through the base pane, use one of the following methods:**

- Drag the view finder
- Use the base pane scroll bars
- Press the left or right arrow key (moves the view one base at a time left or right)

- Press the Page Up or Page Down key (moves the view finder to the next group of sequence maps above or below)

The sample base calls are highlighted with different colors for easy identification. The default colors are:

|               |                                                         |
|---------------|---------------------------------------------------------|
| <b>White</b>  | The sample base call is the same as the reference base. |
| <b>Blue</b>   | A no call (n).                                          |
| <b>Green</b>  | A homozygous difference call.                           |
| <b>Orange</b> | A heterozygous difference call.                         |

## Printing the Sequence Maps

To print the map pane:

1. Click the **Print** button ; or

Select **File** → **Print** from the menu bar.

The Print dialog box opens (see [Figure 4.17 on page 39](#)).

The Print dialog box includes the standard Windows print options (select printer, change properties, number of copies, etc., depending upon your system and printer).

The Selection radio button is disabled in Sequence view.

2. Click **OK**.

The map pane is printed.

## Sequence View Options

In the sequence view, you can modify the:

- Font size in the base pane
- Colors in the map and base panes



**NOTE:** You can also control the display of the genomic positions and PCR start/stop positions; for more information, see [Displaying Genomic and PCR Start/Stop Positions on page 48](#).

---

To change the display of features in the sequence view:

1. Click the **Options** button .

The Resequencing Analysis Options dialog box opens ([Figure 4.24](#)).

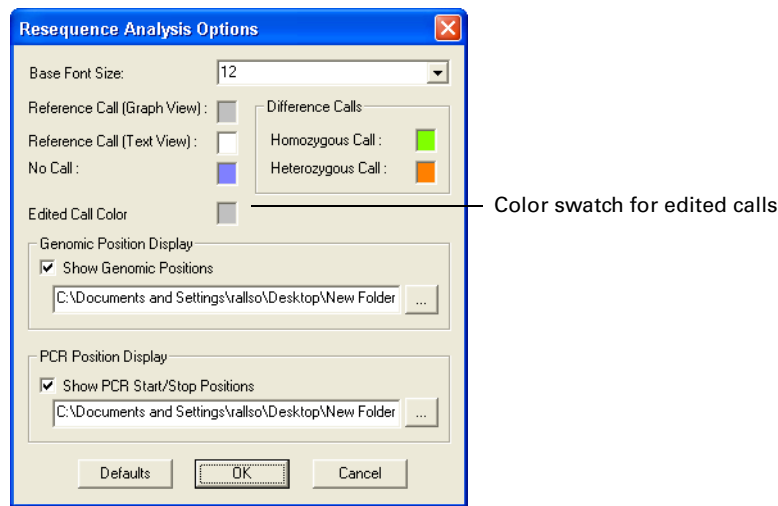


**Figure 4.24** Resequencing Analysis Options dialog box

2. To change the font size for base calls in the base pane, click the **Base Font Size** drop-down arrow and make a selection from the drop-down list.
3. To change the colors in the map or base pane:

A. Click the color swatch (Figure 4.24) for the type of call you want to change:

- Reference Call (Graph View)
- Reference Call (Text View)
- No Call
- Edited Call
- Homozygous Difference Call
- Heterozygous Difference Call

The color palette opens (Figure 4.25).

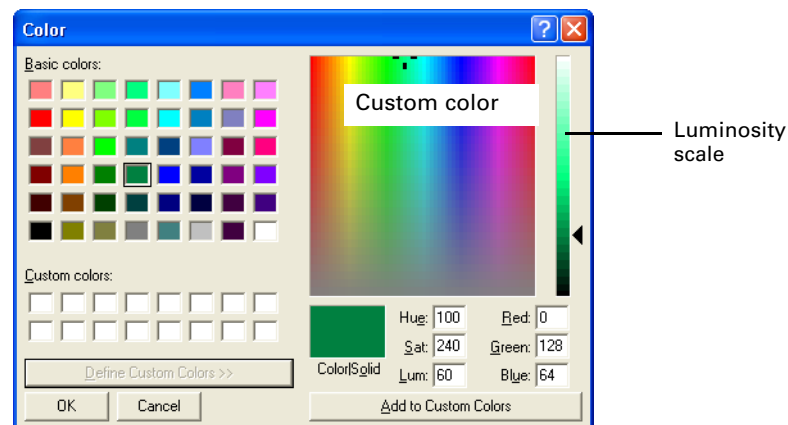


**Figure 4.25** Color palette, custom color field displayed

- B. To select a predefined color, click a basic color.
- C. To define a custom color, click **Define Custom Colors** and use the click-and-drag method to move the cross hairs in the custom color field. Adjust the color brightness using the luminosity scale to the right. When finished, click **Add to Custom Colors** to apply the color.

- D. Click **OK** to close the Color palette.
4. Click **OK** in the Resequencing Analysis Options dialog box.
- The new color is applied to the selected item in the map and base windowpanes.

SNP View

The SNP viewer is similar to the sequence viewer and displays the base calls for bases with one or more SNPs across all samples.

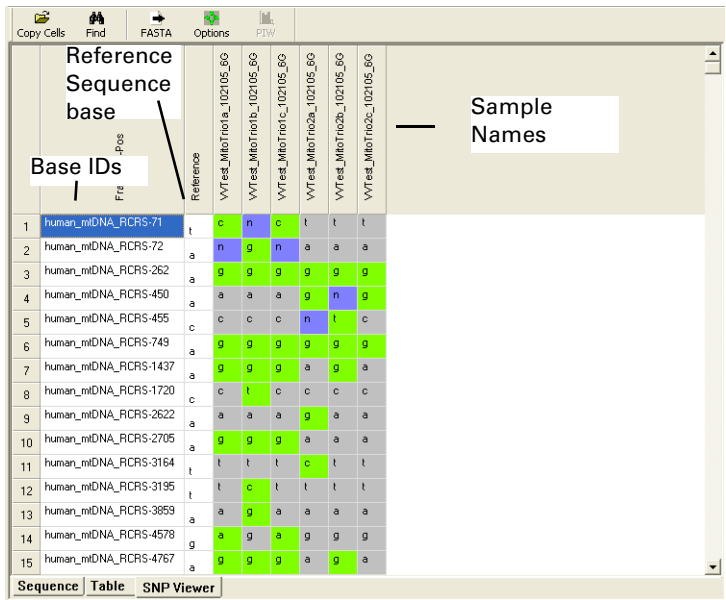


Figure 4.26 SNP Viewer

The viewer displays the following information:

- Fragment ID and Position** The fragment name and sequence position of the reference base in the reference fragment.
- Reference Sequence base** The reference base at the fragment position.
- Sample Names** The analysis results with the SNP.
- SNP calls** The base call at the position.

The viewer uses the following color code to display base calls:

- Gray** The sample base call is the same as the reference base.
- Blue** A no call.
- Green** A homozygous difference call.
- Orange** A heterozygous difference call.

Displaying Genomic and PCR Start/Stop Positions

GSEQ allows you to display the genomic position and the PCR start/stop positions in the Table and Sequence Views of the Resequencing Analysis window.

To do this, you must:

- Create data files with the genomic position and PCR start/stop position data for the array (see below)

- Select the correct data files for the array (see page 50)

The Table view can display the genomic position (see [Table View on page 30](#)).

The Sequence view can display the genomic position and PCR start/stop positions (see [Sequence View on page 40](#)).

## Data File Formats

You must provide the genomic position and PCR start/stop position data in space-delimited text file using the following formats:

### Genomic position file

The genomic position file ([Figure 4.27](#)) for each array type must have a unique name, and must contain the following information for each fragment:

- Fragment Name
- Chromosome in which the fragment resides
- Chromosome position for the first base in the fragment.

|    | A                 | B  | C      |
|----|-------------------|----|--------|
| 1  | AFFX-TagIQ-EX     | 22 | 260000 |
| 2  | DCN_002           | X  | 20000  |
| 3  | DCN_003-4         | X  | 40000  |
| 4  | DCN_005-8         | X  | 60000  |
| 5  | DCN_009           | X  | 80000  |
| 6  | DCN_010           | X  | 100000 |
| 7  | DCN_011-13        | X  | 120000 |
| 8  | DCN_014           | X  | 140000 |
| 9  | DCN_016           | X  | 160000 |
| 10 | DCN_018-19        | X  | 180000 |
| 11 | DCN_020           | X  | 200000 |
| 12 | DCN_021-23        | X  | 220000 |
| 13 | DCN_024-26        | X  | 240000 |
| 14 | DCN_027           | X  | 260000 |
| 15 | DCN_028-29        | X  | 280000 |
| 16 | DCN_030           | X  | 300000 |
| 17 | DCN_031           | X  | 320000 |
| 18 | DCN_032-34        | X  | 340000 |
| 19 | DCN_035           | X  | 360000 |
| 20 | DCN_036           | X  | 380000 |
| 21 | DCN_037-38        | X  | 400000 |
| 22 | DCN_039           | X  | 420000 |
| 23 | DCN_040           | X  | 440000 |
| 24 | DCN_042-43        | X  | 460000 |
| 25 | DCN_044           | 22 | 480000 |
| 26 | DCN_045-46        | 22 | 500000 |
| 27 | DCN_047-49        | 22 | 520000 |
| 28 | DCN_050           | 22 | 540000 |
| 29 | DCN_051           | 22 | 560000 |
| 30 | DCN_052           | 22 | 580000 |
| 31 | DCN_053           | 22 | 600000 |
| 32 | DCN_055           | 22 | 620000 |
| 33 | DCN_058-b         | 22 | 640000 |
| 34 | 011671_p1-6_D10.1 | 22 | 660000 |
| 35 | 011671_p1-6_D10.2 | 22 | 680000 |
| 36 | 011671_p1-6_D10.3 | 22 | 700000 |
| 37 | 011671_p1-6_D10.4 | 22 | 720000 |

**Figure 4.27** Genomic position file

### PCR Start/Stop Position File



**NOTE:** You must provide genomic position data to display PCR start/stop data.

The PCR start/stop position file ([Figure 4.28](#)) for each array type must have a unique name, and must contain the following information for each fragment:

- Tiled Fragment name for the PCR sequence

- Chromosome position for the start position of the PCR sequence
- Chromosome position for the stop position of the PCR sequence

|    | A     | B      | C      |  |
|----|-------|--------|--------|--|
| 1  | PCR0  | 260300 | 260700 |  |
| 2  | PCR1  | 260500 | 300850 |  |
| 3  | PCR2  | 20000  | 60000  |  |
| 4  | PCR3  | 30000  | 90000  |  |
| 5  | PCR4  | 40000  | 120000 |  |
| 6  | PCR5  | 50000  | 150000 |  |
| 7  | PCR6  | 60000  | 180000 |  |
| 8  | PCR7  | 70000  | 210000 |  |
| 9  | PCR8  | 80000  | 240000 |  |
| 10 | PCR9  | 90000  | 270000 |  |
| 11 | PCR10 | 100000 | 300000 |  |
| 12 | PCR11 | 110000 | 330000 |  |
| 13 | PCR12 | 120000 | 360000 |  |
| 14 | PCR13 | 130000 | 390000 |  |
| 15 | PCR14 | 140000 | 420000 |  |
| 16 | PCR15 | 150000 | 450000 |  |
| 17 | PCR16 | 160000 | 480000 |  |
| 18 | PCR17 | 170000 | 510000 |  |
| 19 | PCR18 | 180000 | 540000 |  |
| 20 | PCR19 | 190000 | 570000 |  |
| 21 | PCR20 | 200000 | 600000 |  |
| 22 | PCR21 | 210000 | 630000 |  |
| 23 | PCR22 | 220000 | 660000 |  |
| 24 | PCR23 | 230000 | 690000 |  |
| 25 | PCR24 | 240000 | 720000 |  |
| 26 | PCR25 | 250000 | 750000 |  |

Figure 4.28 PCR Start/Stop file

## Selecting the Position Data



**NOTE:** You must provide genomic position data to display PCR start/stop data.

To select genomic position or PCR data for display:

1. Click the **Options** button ; or  
Select **View** → **Options** from the main menu  
The Resequencing Analysis Options dialog box opens (Figure 4.29).

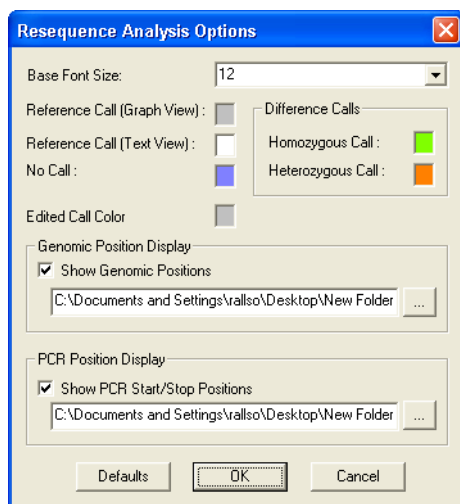


Figure 4.29 Resequencing Analysis Options dialog box

2. Select the checkbox for the desired position display (**Genomic Position** or **PCR Start/Stop Position**).

3. Enter the path and data file name in the textbox; or  
Click the **Browse** button .  
The Select Data dialog box ([Figure 4.30](#)) opens.

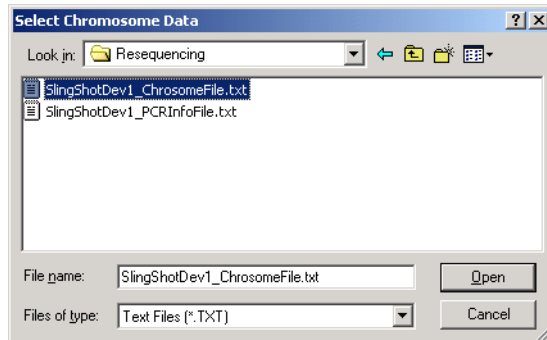


**Figure 4.30** Select Genomic Data dialog box

- A. Use the toolbar to locate the data file.
- B. Select the data file.
- C. Click **Open**.

The path and file name appears in the appropriate position display text box of the Resequencing Analysis Options dialog box.

4. Click **OK**.

The genomic positions are displayed in the:

- Table View (see page 30)
- Sequence View (see page 43)

The PCR positions are only displayed in the [Sequence view](#) (see page 43), and only for those PCR bounds that fall within the fragment.



PROBE INTENSITY WINDOW

The Probe Intensity window (Figure 5.1) displays cell intensity data for Resequencing arrays. You can use the Probe Intensity Window to evaluate the cell intensity data for probe cells used to call a particular feature. This chapter provides an overview of the window and describes the window's:

- [Intensity Plot on page 55](#)
- [Probe Intensity Table on page 58](#)

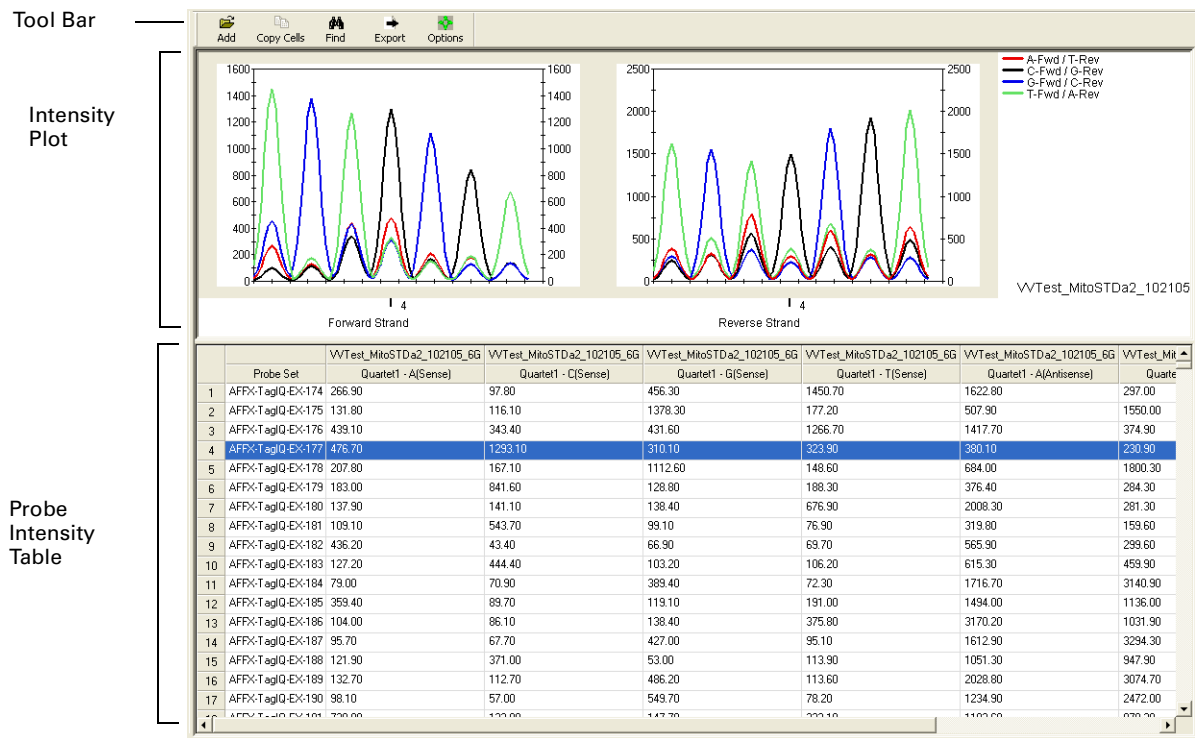


Figure 5.1 Probe Intensity Window for Resequencing array

## Overview

The Probe Intensity Window displays intensity data for the probe cells of selected Resequencing arrays. The Probe Intensity window has two main components:

- [Intensity Plot on page 55](#): For each sample, displays a graphical view of the intensity information for the probe cells of a selected probe set
- [Probe Intensity Table on page 58](#): Displays a list of the probe sets with intensity information for each of the samples. Select a probe set in the Table to display in the Intensity Plot.

For [Resequencing assays](#), the Probe Intensity Window displays intensity information for the probe cells used to call a base in a reference sequence (see page 58).

## Displaying Probe Intensity Data

**To view the probe intensity data for a particular feature in the Resequencing Analysis window:**

1. Select the feature whose intensity values you wish to evaluate.
2. Click the **Probe Intensity Window** button  in the Analysis Results window toolbar.

The Add Cell Intensity Data dialog box opens ([Figure 5.2](#)).

The Add Cell Intensity Data dialog box provides [several options for displaying files](#) (see page 28).

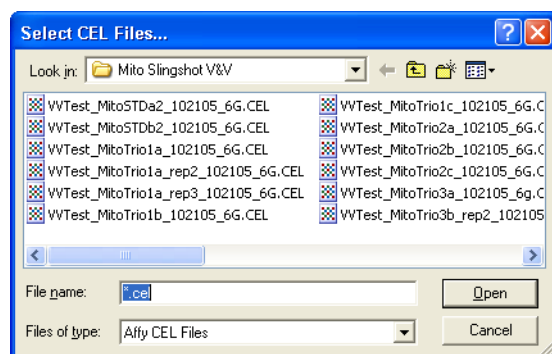


**Figure 5.2** Add Cell Intensity Data dialog box

3. Select the cell intensity data for the sample that you want to view and click **OK**.



**NOTE:** You can select cell intensity data for up to ten samples at a time. If you wish to display data for additional samples, you will have to select them in separate operations.

The Probe Intensity window displays the selected cell intensity data.

The feature selected in the Resequencing Analysis window is automatically selected in the Probe Intensity Table, and the intensity values for that feature are displayed in the Intensity Plot.

**To open the Probe Intensity window and select cell intensity data for display:**

1. Click the **Probe Intensity window** button  in the Tools shortcut bar, or select **Run → Probe Intensity** from the menu bar.

The Add Cell Intensity Data dialog box opens ([Figure 5.2](#)).

The Add Cell Intensity Data dialog box provides [several options for displaying files](#) (see page 28).

2. Select the cell intensity data for the sample that you want to view and click **OK**.

The Probe Intensity window displays the selected cell intensity data.

**To display additional sample intensity data in an open Probe Intensity window:**

1. Click the **Add Cell Intensity data** button  in the Probe Intensity window toolbar, or select **Edit → Add** from the menu bar.

The Add Cell Intensity Data dialog box opens (Figure 5.2).

The Add Cell Intensity Data dialog box provides [several options for displaying files](#) (see page 28).

2. Select the cell intensity data for the sample that you want to view and click **OK**.

The Probe Intensity window displays the selected cell intensity data.

## Probe Intensity Window Toolbar

The tool bar (Figure 5.3) provides quick access to some of the commonly used functions of the Probe Intensity window.



**Figure 5.3** Probe Intensity window toolbar

You can display toolbars with text labels. To display the toolbar button labels, select **View → Toolbar → Text Labels** from the menu bar.

**Table 5.1** Probe Intensity window toolbar button functions

| Menu Bar Command       | Toolbar Button                                                                      | Function                                                                                 |
|------------------------|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| Edit → Add             |  | Opens the <a href="#">Add Cell Intensity Data dialog box</a> (see page 54).              |
| Edit → Copy Item       |  | <a href="#">Copies selected data</a> in the table to the system clipboard (see page 59). |
| Edit → Find            |  | Opens the <a href="#">Find dialog box</a> (see page 59).                                 |
| Export → Table         |  | Opens the <a href="#">Export As dialog box</a> (see page 59).                            |
| Tools → Settings → PIW |  | Displays the <a href="#">Probe Intensity View Options dialog box</a> (see page 57).      |

## Intensity Plot

The Intensity Plot displays the intensity information for the probe cells of the selected feature as a graph (Figure 5.4).

**To select a probe set to display in the Intensity Plot:**

- Click anywhere in the probe set's row in the Probe Intensity table.

You can select from the following formats:

- Line Graph or Bar Graph (see below)
- Trace Plot (see page 57)

## Bar and Line Graph

For Resequencing assays (Figure 5.4) the plot displays the intensity values for the probe cells used for the selected fragment and reference base (see page 58).

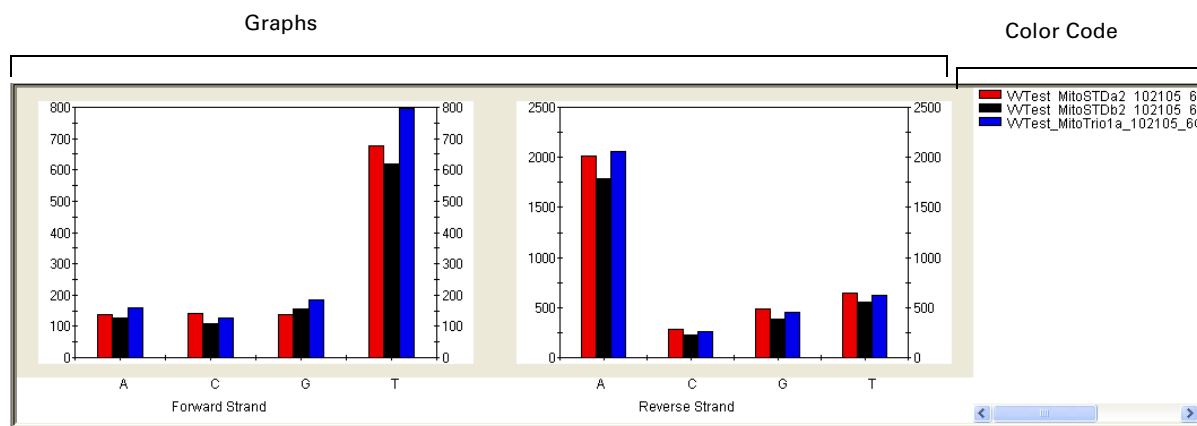


**Figure 5.4** Intensity Plot for GSEQ assay, bar graph format

The left side of the graph plot displays the graphs of intensity data for the selected feature. The intensity

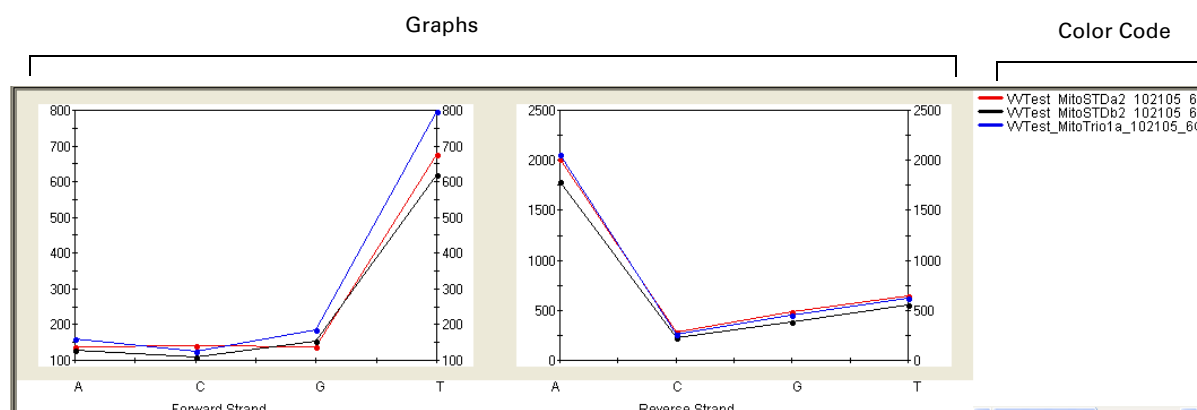


**Figure 5.5** Intensity Plot for GSEQ assay, line graph format

values for each sample are represented by lines or bars of different colors. Appropriate scale ranges are selected for the intensity values.

The right side provides a color code to distinguish the intensity values from different samples.

### Changing Display Options for the Line and Bar Graphs

You can change between the Line and Bar Graph displays:

**To change the graph format:**

1. Click the **Options** button .

The Probe Intensity View Options dialog box opens (Figure 5.6).

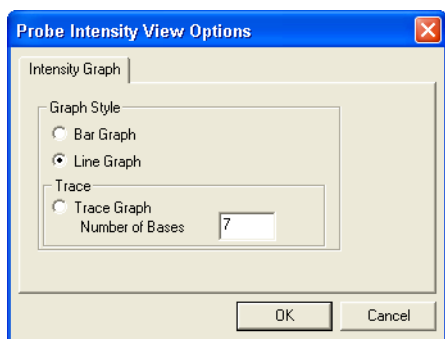


**Figure 5.6** Probe Intensity View Options dialog box

2. Select the desired display option and click **OK**.  
The desired display opens in the Probe Intensity window.

## Trace Plot

The trace view displays intensities of neighboring bases in the same chart.

Plots from different samples are stacked vertically in the plot.

A Gaussian curve is used to represent the intensity of each probe. The height of the curve represents the probe intensity for the probe. Probes from neighboring bases are displayed sequentially in the chart along the X axis. Each base type (A, C, T, and G) has drawn with a different color.

The probe intensity for forward strand and reverse strand probe quartets are drawn in adjacent charts.

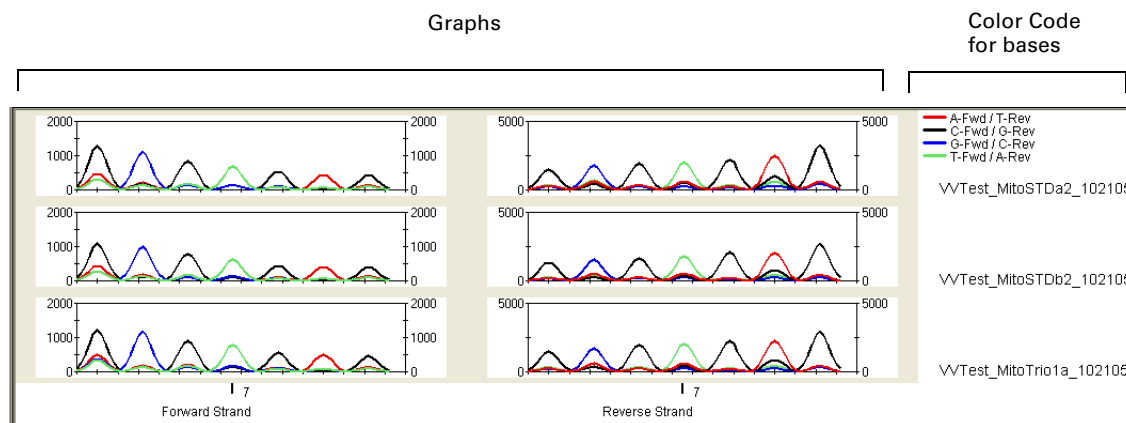


**Figure 5.7** Trace Plot for GSEQ assay

## Changing Display Options for the Trace Plot

To select the trace plot and change the number of bases displayed:

1. Click the **Options** button .  
The Probe Intensity View Options dialog box opens (Figure 5.8).

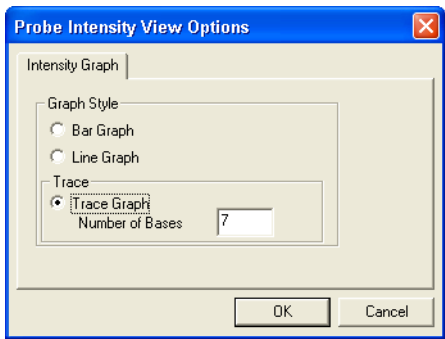


Figure 5.8 Probe Intensity View Options dialog box

- 2. Select the Trace Graph option.
- 3. Enter a number between 1 and 100 for the Number of bases box.
- 4. Click OK.

The desired display opens in the Probe Intensity window.

Probe Intensity Table

The Probe Intensity table displays a list of the probe sets used in the selected arrays. The first column in each row displays the probe set ID. The rest of the columns display intensity information for each selected sample's cell intensity data file; for each sample, the table cells in a row display the individual cell intensity data for the probe cells of the probe set. The number of probe cells displayed for each sample depends upon the assay and probe array type.

You can:

- Search the data in the table (see page 59)
- Sort the table (see page 59)
- Export data (see page 59)

Data Displayed for Resequencing Arrays

Probe Set ID

|    |                   | VVTest_MitoSTDa2_102105_6G | VVTest_MitoSTDa2_102105_6G | VVTest_MitoSTDa2_102105_6G | VVTest_MitoSTDa2_102105_6G | VV | Sample Name   |
|----|-------------------|----------------------------|----------------------------|----------------------------|----------------------------|----|---------------|
|    | Probe Set         | Quartet1 - A(Sense)        | Quartet1 - C(Sense)        | Quartet1 - G(Sense)        | Quartet1 - T(Sense)        |    | Probe Cell ID |
| 16 | AFFX-TagIQ-EX-189 | 132.70                     | 112.70                     | 486.20                     | 113.60                     | 20 |               |
| 17 | AFFX-TagIQ-EX-190 | 98.10                      | 57.00                      | 549.70                     | 78.20                      | 12 |               |
| 18 | AFFX-TagIQ-EX-191 | 728.90                     | 123.80                     | 147.70                     | 232.10                     | 11 |               |
| 19 | AFFX-TagIQ-EX-192 | 182.20                     | 107.00                     | 220.40                     | 794.10                     | 24 |               |
| 20 | AFFX-TagIQ-EX-193 | 142.00                     | 74.90                      | 165.70                     | 896.20                     | 23 |               |
| 21 | AFFX-TagIQ-EX-194 | 165.20                     | 106.30                     | 269.20                     | 908.60                     | 20 |               |
| 22 | AFFX-TagIQ-EX-195 | 67.80                      | 59.70                      | 837.20                     | 62.20                      | 61 |               |
| 23 | AFFX-TagIQ-EX-196 | 686.70                     | 78.40                      | 101.80                     | 157.60                     | 34 |               |
| 24 | AFFX-TagIQ-EX-197 | 311.30                     | 227.40                     | 120.00                     | 140.00                     | 27 |               |

Figure 5.9 Probe Intensity Table for Resequencing arrays

For Resequencing arrays, the Probe Intensity Table displays a list of the fragment positions used in the arrays, with the following information presented for each fragment position and sample (Figure 5.9):

|              |                                                                                     |                         |
|--------------|-------------------------------------------------------------------------------------|-------------------------|
| Probe Set ID | Unique identifier for the Fragment, with the following information for each sample: |                         |
| Sample Name  | With assay results:                                                                 |                         |
|              | Sense Probe Quartet                                                                 | Antisense Probe Quartet |
| A            | Intensity Value                                                                     | Intensity Value         |
| C            | Intensity Value                                                                     | Intensity Value         |
| T            | Intensity Value                                                                     | Intensity Value         |
| G            | Intensity Value                                                                     | Intensity Value         |

Since a sample has two pairs of probe quartets, and the table can display more than one sample, you will need to use the scroll bars to view all the data in the table.

## Searching the Table

You can perform a text search on the cells in the table using the Find dialog box.

**To open the Find dialog box:**

- Click the **Find** button , or click **Edit** → **Find** in the menu bar.

## Sorting the Table

### Single Column Table Sort

You can sort the table on any data column in ascending or descending order.

- In the table, right-click the column header of interest and click **Sort Ascending** or **Sort Descending** in the shortcut menu.

### Multiple Column Table Sort

You can sequentially sort the table using up to four data columns using the Sort dialog box.

**To open the Sort dialog box:**

- Right-click the table and select **Sort** from the shortcut menu, or select **Edit** → **Sort** from the menu bar.

For more information, see [Multiple Column Table Sort on page 32](#).

## Exporting Cell Intensity Data

You can export data from the feature table for further analysis in other applications by:

- Exporting data as a tabular format file (see page 35)
- Copying selected cells to the system Clipboard (see page 36)



## Chapter 6

# RESEQUENCING ALGORITHM REPORTS

After running a resequencing analysis, GSEQ automatically generates an algorithm report. The report includes a variety of information about the analyses, including:

- A list of the samples analyzed
- A summary of information about the base calls
- Quality control information

This chapter describes the contents of the Resequencing Algorithm Report (see below).

## Resequencing Algorithm Report Contents

**To open a report:**

1. From the **File** menu, select **Open Report...**

The Open Report dialog box opens (Figure 6.1).

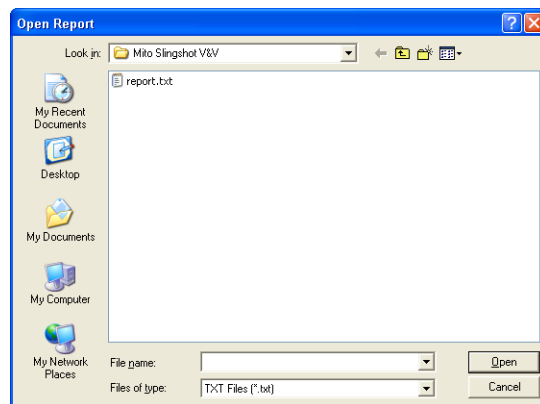


**Figure 6.1** Open dialog box

2. In the **Files of type** drop-down list, choose **Reports (\*.rpt)**.
3. Double-click the report you want to view, or select multiple reports and click **Open**.  
The report is displayed in Notepad.



**CAUTION:** You can also open other data files in Notepad from this dialog box. If you open a data file and inadvertently edit and save it in Notepad you may render it unreadable in GSEQ and AGCC.

For more information, see [Chapter 6, Resequencing Algorithm Reports](#) on page 61.



**TIP:** You can also display report files in Excel or other spreadsheet software.

Heading

Array data and Algorithm Settings

Sample Summary

Call Rate Summary

| report.txt |                                                         |                                             |                 |                 |                 |                 |                 |          |  |
|------------|---------------------------------------------------------|---------------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|----------|--|
| 1          | Resequencing Algorithm Report                           | algorithm: Reseq1                           | version: 2      |                 |                 |                 |                 |          |  |
| 2          | Report File Name -                                      | C:\Command_Console\Data\Reseq_03\report.txt |                 |                 |                 |                 |                 |          |  |
| 3          | Date:                                                   | 1/4/2008 14:03                              |                 |                 |                 |                 |                 |          |  |
| 4          | Probe Array Type -                                      | DCNtaglQr510989                             |                 |                 |                 |                 |                 |          |  |
| 5          |                                                         |                                             |                 |                 |                 |                 |                 |          |  |
| 6          | Number of Bases per Array                               | 27670                                       |                 |                 |                 |                 |                 |          |  |
| 7          | Number of Samples                                       | 8                                           |                 |                 |                 |                 |                 |          |  |
| 8          | Parameters Used By Algorithm:                           |                                             |                 |                 |                 |                 |                 |          |  |
| 9          | No Signal Threshold (probe signal/noise ratio) = 1      |                                             |                 |                 |                 |                 |                 |          |  |
| 10         | Weak Signal Fold Threshold (mean/probe ratio) = 20      |                                             |                 |                 |                 |                 |                 |          |  |
| 11         | Large SNR Threshold (probe signal/noise ratio) = 20     |                                             |                 |                 |                 |                 |                 |          |  |
| 12         | Strand Quality Threshold (quality score) = 0            |                                             |                 |                 |                 |                 |                 |          |  |
| 13         | Total Quality Threshold (quality score) = 75            |                                             |                 |                 |                 |                 |                 |          |  |
| 14         | Maximum Fraction of Heterozygote Calls (0-1) = 0.9      |                                             |                 |                 |                 |                 |                 |          |  |
| 15         | Model Type (0=Heterozygote, 1=Homozygote) = 0           |                                             |                 |                 |                 |                 |                 |          |  |
| 16         | Perfect Call Quality Threshold (quality score) = 2      |                                             |                 |                 |                 |                 |                 |          |  |
| 17         | Min Fraction of Calls in Neighboring Probes (0-1) = 0.5 |                                             |                 |                 |                 |                 |                 |          |  |
| 18         | Min Fraction of Calls of Samples (0-1) = 0.75           |                                             |                 |                 |                 |                 |                 |          |  |
| 19         |                                                         |                                             |                 |                 |                 |                 |                 |          |  |
| 20         | Summary:                                                |                                             |                 |                 |                 |                 |                 |          |  |
| 21         |                                                         |                                             |                 |                 |                 |                 |                 |          |  |
| 22         | Sample                                                  | Total                                       | Called          | CallRate        | NS              | WS              | SAT             | FinalRel |  |
| 23         | 5303_DCN_01                                             | 27670                                       | 26190           | 94.65%          | 6               | 1               | 0               |          |  |
| 24         | 5303_DCN_02                                             | 27670                                       | 25320           | 91.51%          | 15              | 3               | 0               |          |  |
| 25         | 5303_DCN_03                                             | 27670                                       | 26266           | 95.00%          | 6               | 0               | 0               |          |  |
| 26         | 5303_DCN_04                                             | 27670                                       | 26181           | 94.62%          | 5               | 2               | 0               |          |  |
| 27         | 5303_DCN_05                                             | 27670                                       | 26418           | 95.48%          | 9               | 0               | 0               |          |  |
| 28         | 5303_DCN_06                                             | 27670                                       | 26142           | 94.48%          | 1               | 1               | 0               |          |  |
| 29         | 5303_DCN_07                                             | 27670                                       | 26064           | 94.20%          | 8               | 0               | 0               |          |  |
| 30         | 5303_DCN_08                                             | 27670                                       | 26345           | 95.21%          | 12              | 0               | 0               |          |  |
| 31         |                                                         |                                             |                 |                 |                 |                 |                 |          |  |
| 32         | Call Rate Summary                                       |                                             |                 |                 |                 |                 |                 |          |  |
| 33         |                                                         |                                             | New_5303_DCN_01 | New_5303_DCN_01 | New_5303_DCN_02 | New_5303_DCN_02 | New_5303_DCN_03 | New_5303 |  |
| 34         | Fragment                                                | NumBases                                    | CallRate        | RefCalls        | CallRate        | RefCalls        | CallRate        | RefCalls |  |
| 35         | IO_EX                                                   | 4086                                        | 88.40%          | 99.90%          | 86.20%          | 99.90%          | 88.20%          |          |  |
| 36         | DCN002                                                  | 252                                         | 98.00%          | 100.00%         | 98.00%          | 100.00%         | 98.00%          |          |  |
| 37         | DCN003-4                                                | 1843                                        | 97.10%          | 100.00%         | 94.20%          | 100.00%         | 97.50%          |          |  |
| 38         | DCN005-8                                                | 2253                                        | 95.80%          | 100.00%         | 93.40%          | 100.00%         | 96.50%          |          |  |
| 39         | DCN009                                                  | 282                                         | 95.40%          | 100.00%         | 88.30%          | 100.00%         | 96.10%          |          |  |
| 40         | DCN010                                                  | 2341                                        | 96.40%          | 100.00%         | 94.30%          | 99.90%          | 96.60%          |          |  |
| 41         | DCN011-13                                               | 1275                                        | 94.80%          | 100.00%         | 88.40%          | 99.90%          | 95.50%          |          |  |
| 42         | DCN014                                                  | 135                                         | 97.00%          | 100.00%         | 94.10%          | 100.00%         | 98.50%          |          |  |

**Figure 6.2** Resequencing Algorithm Version 2 report (displayed in Excel)

The report summarizes information about the analysis of the samples and the arrays used ([Figure 6.2](#)).

The report has the following components:

- [Heading](#)
- [Array Data and Algorithm Parameters](#) on page 63
- [Sample Summary](#) on page 64
- [Fragment Call Rate Summary](#) on page 64

**Heading**

|   |                               |                                             |            |
|---|-------------------------------|---------------------------------------------|------------|
| 1 | Resequencing Algorithm Report | algorithm: Reseq1                           | version: 2 |
| 2 | Report File Name -            | C:\Command_Console\Data\Reseq_03\report.txt |            |
| 3 | Date:                         | 1/4/2008 14:03                              |            |
| 4 | Probe Array Type -            | DCNtaglQr510989                             |            |

**Figure 6.3** Resequencing Algorithm Report: Heading

The report heading ([Figure 6.3](#)) provides basic information about the report:

Report type

Report File Name

Date Date report was generated.

Probe Array Type

## Array Data and Algorithm Parameters

This section ( [Figure 6.4](#), [Figure 6.5](#))) provides basic information about the array and the algorithm parameters used in the analysis:

**Number of Bases per Array** The number of bases interrogated by the Resequencing probe array.

**Number of Samples** The number of samples in the analysis.

**Parameters used by Algorithm** Algorithm parameters set by user.

The appropriate parameters are displayed for the algorithm type. For more information about the algorithm parameters, see:

- [Resequencing Algorithm Version 1 Settings on page 74](#) ([Figure 6.4](#))

|    |                                                         |       |  |
|----|---------------------------------------------------------|-------|--|
| 5  |                                                         |       |  |
| 6  | Number of Bases per Array                               | 27670 |  |
| 7  | Number of Samples                                       | 8     |  |
| 8  | Parameters Used By Algorithm:                           |       |  |
| 9  | No Signal Threshold (probe signal/noise ratio) = 1      |       |  |
| 10 | Weak Signal Fold Threshold (mean/probe ratio) = 20      |       |  |
| 11 | Large SNR Threshold (probe signal/noise ratio) = 20     |       |  |
| 12 | Strand Quality Threshold (quality score) = 0            |       |  |
| 13 | Total Quality Threshold (quality score) = 75            |       |  |
| 14 | Maximum Fraction of Heterozygote Calls (0-1) = 0.9      |       |  |
| 15 | Model Type (0=Heterozygote, 1=Homozygote) = 0           |       |  |
| 16 | Perfect Call Quality Threshold (quality score) = 2      |       |  |
| 17 | Min Fraction of Calls in Neighboring Probes (0-1) = 0.5 |       |  |
| 18 | Min Fraction of Calls of Samples (0-1) = 0.75           |       |  |

**Figure 6.4** Resequencing Algorithm Version 1 Report: Array data and algorithm parameters

- [Resequencing Algorithm Version 2 Settings on page 82](#) ([Figure 6.5](#))

|    |                                                     |       |  |  |
|----|-----------------------------------------------------|-------|--|--|
| 5  |                                                     |       |  |  |
| 6  | Number of                                           | 37756 |  |  |
| 7  | Number of                                           | 14    |  |  |
| 8  | Parameters Used By Algorithm:                       |       |  |  |
| 9  | No Signal Threshold (probe signal/noise ratio) = 2  |       |  |  |
| 10 | Weak Signal Fold Threshold (mean/probe ratio) = 20  |       |  |  |
| 11 | Large SNR Threshold (probe signal/noise ratio) = 20 |       |  |  |
| 12 | Perfect Call Quality Threshold (quality score) = 12 |       |  |  |
| 13 | Model Type (0=Heterozygote, 1=Homozygote) = 1       |       |  |  |
| 14 | Min Fraction of Calls of Samples (0-1) = 0.5        |       |  |  |
| 15 | Trace Threshold = 1                                 |       |  |  |
| 16 | Sequence Profile Threshold = -0.175                 |       |  |  |

**Figure 6.5** Resequencing Algorithm Version 2 Report: Array data and algorithm parameters

## Sample Summary

|    |             |       |        |          |    |    |     |          |
|----|-------------|-------|--------|----------|----|----|-----|----------|
| 20 | Summary:    |       |        |          |    |    |     |          |
| 21 |             |       |        |          |    |    |     |          |
| 22 | Sample      | Total | Called | CallRate | NS | WS | SAT | FinalRel |
| 23 | 5303_DCN_01 | 27670 | 26190  | 94.65%   | 6  | 1  | 0   | 55       |
| 24 | 5303_DCN_02 | 27670 | 25320  | 91.51%   | 15 | 3  | 0   | 65       |
| 25 | 5303_DCN_03 | 27670 | 26286  | 95.00%   | 6  | 0  | 0   | 49       |
| 26 | 5303_DCN_04 | 27670 | 26181  | 94.62%   | 5  | 2  | 0   | 35       |
| 27 | 5303_DCN_05 | 27670 | 26418  | 95.48%   | 9  | 0  | 0   | 62       |
| 28 | 5303_DCN_06 | 27670 | 26142  | 94.48%   | 1  | 1  | 0   | 28       |
| 29 | 5303_DCN_07 | 27670 | 26064  | 94.20%   | 8  | 0  | 0   | 126      |
| 30 | 5303_DCN_08 | 27670 | 26345  | 95.21%   | 12 | 0  | 0   | 64       |

**Figure 6.6** Resequencing Algorithm Report: Sample Summary

The Sample Summary (Figure 6.6) provides a sample-by-sample breakdown of the performance:

|                                     |                                                                                                                                                                                                                                                                 |
|-------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Sample</b>                       | An identifier assigned to the analyzed samples.                                                                                                                                                                                                                 |
| <b>Total</b>                        | Total number of bases per sample.                                                                                                                                                                                                                               |
| <b>Called</b>                       | The number of bases called for the sample.                                                                                                                                                                                                                      |
| <b>CallRate</b>                     | Calls/Number of bases per array.                                                                                                                                                                                                                                |
| <b>NS (No Signal)</b>               | The number of sites (base positions) ruled a no call because a feature for the site has a mean intensity within two standard deviations of zero                                                                                                                 |
| <b>WS (Weak Signal)</b>             | The number of sites (base positions) ruled a no call because the highest mean intensity of a feature for the sense or antisense strand was 20-fold lower than the highest mean intensity feature averaged across all samples for the sense or antisense strand. |
| <b>Sat (Saturation)</b>             | The number of sites (base positions) ruled a no call because two probe cells (haploid data) or three probe cells (diploid data) in a four probe cell set for the sense or antisense strand were within two standard deviations of the saturation level.         |
| <b>FinalRel (Final Reliability)</b> | The number of bases whose calls were changed to no call by the final reliability rules.                                                                                                                                                                         |

## Fragment Call Rate Summary

|    |                   |          |                 |                 |                 |                 |                 |
|----|-------------------|----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 32 | Call Rate Summary |          |                 |                 |                 |                 |                 |
| 33 |                   |          | New_5303_DCN_01 | New_5303_DCN_01 | New_5303_DCN_02 | New_5303_DCN_02 | New_5303_DCN_03 |
| 34 | Fragment          | NumBases | CallRate        | RefCalls        | CallRate        | RefCalls        | CallRate        |
| 35 | IQ_EX             | 4086     | 88.40%          | 99.90%          | 86.20%          | 99.90%          | 88.20%          |
| 36 | DCN002            | 252      | 98.00%          | 100.00%         | 98.00%          | 100.00%         | 98.00%          |
| 37 | DCN003-4          | 1843     | 97.10%          | 100.00%         | 94.20%          | 100.00%         | 97.50%          |
| 38 | DCN005-8          | 2253     | 95.80%          | 100.00%         | 93.40%          | 100.00%         | 96.50%          |
| 39 | DCN009            | 282      | 95.40%          | 100.00%         | 88.30%          | 100.00%         | 96.10%          |
| 40 | DCN010            | 2341     | 96.40%          | 100.00%         | 94.30%          | 99.90%          | 96.60%          |
| 41 | DCN011-13         | 1275     | 94.80%          | 100.00%         | 88.40%          | 99.90%          | 95.50%          |
| 42 | DCN014            | 135      | 97.00%          | 100.00%         | 94.10%          | 100.00%         | 98.50%          |
| 43 | DCN016            | 984      | 97.80%          | 100.00%         | 96.40%          | 99.90%          | 97.70%          |
| 44 | DCN018-19         | 364      | 94.80%          | 100.00%         | 93.10%          | 100.00%         | 95.10%          |
| 45 | DCN020            | 162      | 98.80%          | 100.00%         | 98.10%          | 100.00%         | 98.80%          |
| 46 | DCN021-23         | 1126     | 96.30%          | 100.00%         | 89.60%          | 100.00%         | 96.50%          |
| 47 | DCN024-26         | 1424     | 97.30%          | 99.90%          | 92.20%          | 100.00%         | 97.30%          |
| 48 | DCN027            | 469      | 95.90%          | 100.00%         | 87.40%          | 100.00%         | 97.00%          |
| 49 | DCN028-29         | 581      | 96.60%          | 100.00%         | 96.60%          | 100.00%         | 96.60%          |

**Figure 6.7** Resequencing Algorithm Report: Fragment Call Rate Summary

The summary (Figure 6.7) provides a breakdown of the performance by fragment and sample:

**Fragment** ID assigned to the fragment.

**NumBases** Number of bases in the fragment.

**For each sample:**

- **Sample ID**
- **Call Rate** Calls/Number of bases per fragment.
- **RefCalls** Percentage of calls in fragment that matched reference sequence.



# INSTALLING GSEQ

This appendix provides detailed instructions for installing the GeneChip® Sequence Analysis Software (GSEQ). It includes the following sections:

- [Requirements](#)
- [Installation on page 67](#)

## Requirements



**NOTE:** The screen captures depicted in this manual may not exactly match the windows displayed on your screen.

It is recommended that there is at least 500 MB of available disk space for the installation.

GSEQ 4.1 can be installed:

- On a computer with no other AGCC software installed
- On a computer with GCOS installed
- On a computer with AGCC installed

## Installation

**To install the GSEQ software:**

1. Download the GeneChip® Sequence Analysis Software 4.1 from the Affymetrix Web Site to a directory on your computer.
2. Unzip the installation files to another directory on your computer.
3. Click Install.exe in that directory.

The GSEQ splash screen opens ([Figure A.1](#)), followed by the Install Welcome window ([Figure A.2](#)).

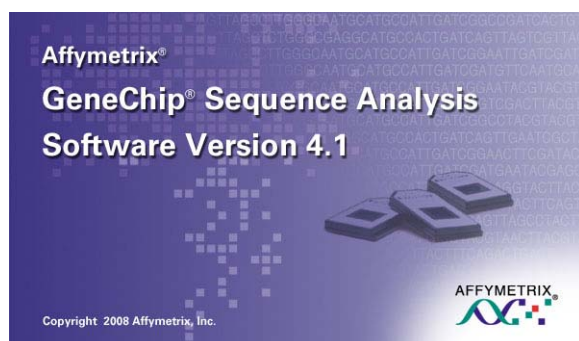


Figure A.1 GSEQ Splash Screen

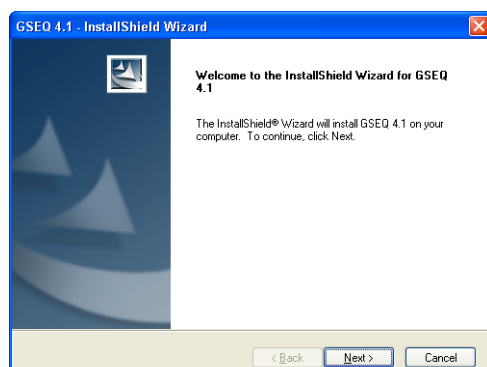


Figure A.2 Install Welcome window (GSEQ)

4. Click **Next**.

The Affymetrix, Inc. End User License Agreement opens (Figure A.3).

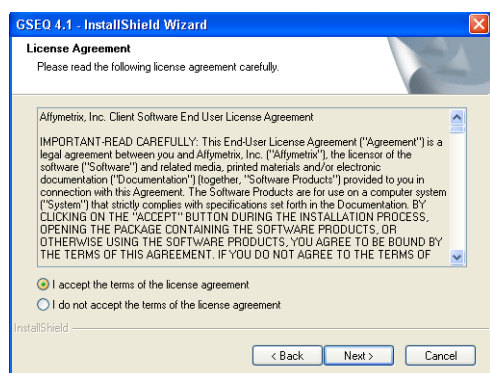


Figure A.3 GSEQ License Agreement window

5. Review the contents and click **Yes** in each window to accept the terms of the licensing agreement. If the **No** option is selected, the installation program will exit and GSEQ will not be installed.

6. Click **Next**.

The **Choose Destination Location** window opens (Figure A.4).

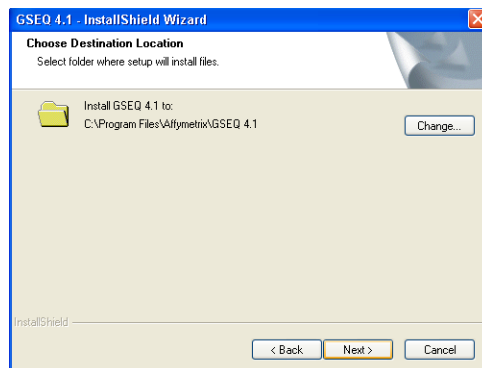


**Figure A.4** GSEQ Choose Destination Location window

7. Select the destination where GSEQ will be installed. C:\Program Files\Affymetrix\GSEQ 4.1 is the default location.
8. Click **Next**.  
The Ready to Install window opens (Figure A.5).

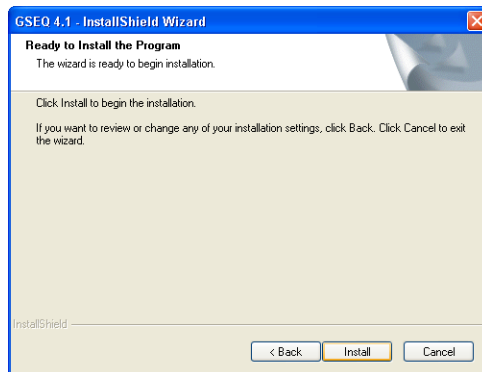


**Figure A.5** GSEQ Ready to Install window

9. Click **Install**.  
The installation begins.  
The Setup Status window opens (Figure A.6).

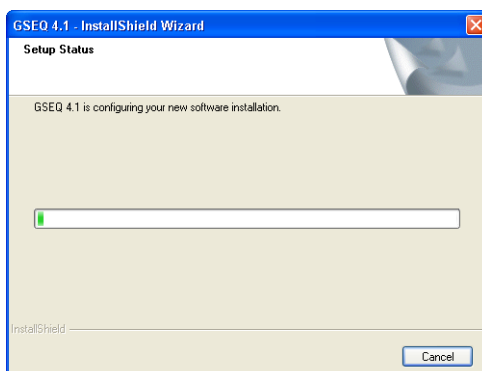


**Figure A.6** GSEQ Setup Status window

After the files are copied, the Install Complete window open ([Figure A.7](#)).

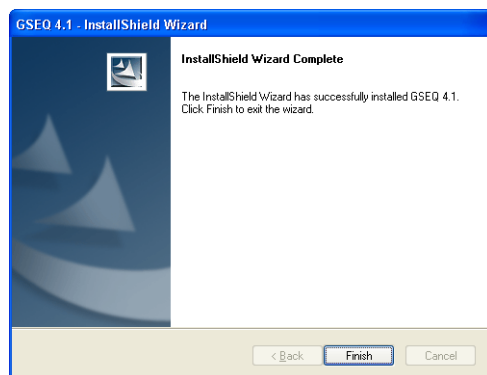


**Figure A.7** GSEQ Install Complete window

**10.** To exit the InstallShield Wizard, click **Finish**.

## RESEQUENCING ALGORITHM VERSION 1

The Resequencing algorithm analyzes the cell intensity data (\*.cel) from an Affymetrix® resequencing probe array.

This appendix includes:

- [Resequencing Algorithm Version 1 Description](#)
- [Resequencing Algorithm Version 1 Settings on page 74](#)

### Resequencing Algorithm Version 1 Description

This section contains information on the following topics:

- The notation used for the resequencing probe array (see page 71)
- The data filters (see page 71)
- The DNA analysis models (see page 72)
- The final reliability rules (see page 73)
- Heterozygosity calculations (see page 74)
- SNP calculations (see page 74)
- References (see page 74)

### Notation

An Affymetrix resequencing probe array consists of a number of probe cells or features (20x24  $\mu\text{m}$  rectangular areas on the array). During scanning, the software divides each probe cell into subunit squares or pixels (3x3  $\mu\text{m}$ ). Each probe cell contains many copies of a unique 25-base oligonucleotide probe of defined sequence. Eight probe cells query a specific site in a known reference sequence. Four probe cells interrogate the sense strand and contain probes that are identical except for the central base which is A, C, G, or T. Four probe cells interrogate the antisense strand and contain probes that are identical except for the central base which is A, C, G, or T.

The Affymetrix GeneChip® Command Console (AGCC) computes the mean and standard deviation for a data file (\*.dat) and saves these to a cell intensity file (\*.cel).

### Data Filters

The data filters identify unreliable data or adjust the extremely small variance of intensities that approach the detector limits.



**NOTE:** The filter thresholds and other parameters can be adjusted in the Reseq Algorithms dialog box. See [Resequencing Algorithm Version 1 Settings on page 74](#) for more information.

### No Signal

If any feature for a site (in the sense or antisense strand) has a mean intensity within two standard deviations of zero, the site is ruled a no call (n) for the sample.

### Weak Signal

If the highest mean intensity of a feature for the sense (or antisense) strand is 20-fold lower than the highest mean intensity feature, averaged across all samples, for the sense (or antisense) strand, the site is ruled n for the sample.

### Saturation

If two probe cells (haploid data) or three probe cells (diploid data) in a four probe cell set for the sense or antisense strand are within two standard deviations of saturation level, the site is ruled n for the sample.

### Large Signal-to-Noise Ratio

A signal with a very large signal-to-noise ratio (SNR) tends to have an exaggerated influence on the calling algorithm. If the SNR is greater than 20, the variance is set so that the SNR = 20.

## DNA Analysis Models

The models assume the pixel intensities of a feature are independently and normally distributed. The algorithm computes the estimated mean background and variance for the sense and antisense strand features. Initially, the data are analyzed using even background models that assume the same mean background intensity and variance across all features.

The Resequencing calling algorithm is based on the Adaptive Background Genotype Calling Scheme (ABACUS) developed by Cutler et al. (1). This algorithm specifies models for the presence or absence of various genotypes in the sample. For haploid data there are five genotype models (A, C, G, T, no call) and for diploid data there are 11 genotype models (A, C, G, T, AC, AG, AT, CG, CT, GT, no call).

Each model generates an estimate of feature intensity and variation based on a likelihood model. The likelihood of each model is computed as:  $\ln(L) = -\frac{1}{2} \sum_x N_x [\ln(\hat{\sigma}_x^2) + (V_x + M_x^2 - 2\hat{\mu}_x M_x + \hat{\mu}_x^2) / \hat{\sigma}_x^2 + \ln(2\pi)]$

where:

$N_x$  = the number of pixels observed in feature x

$V_x$  = the observed variance for feature x

$M_x$  = observed mean for feature x

$\hat{\mu}_x$  = estimated mean for feature x under the model in question

$\hat{\sigma}_x^2$  = estimated variance for feature x

The sum is taken over all features x, where x is A, C, G, or T, on the sense and antisense strand features. Note that  $\hat{\mu}_x$  and  $\hat{\sigma}_x^2$  differ for each model.

The algorithm computes the likelihood of a particular genotype model independently for the sense strand and antisense strand. The overall likelihood of the model (L) is the product of the likelihood of the sense strand and the likelihood of the antisense strand.

Each genotype model has three quality scores. For example, for model A:

$$Q_f(A) = \log(L_f(A)) - \log(L_f(\text{max other}))$$

$$Q_r(A) = \log(L_r(A)) - \log(L_r(\text{max other}))$$

$$Q(A) = \log(L(A)) - \log(L(\text{max other}))$$

where:

$Q_f(A)$  = sense strand quality score for model A

$Q_r(A)$  = antisense strand quality score for model A

$Q(A)$  = overall or total quality score for model A

$L_f(A)$  = sense strand likelihood of model A

$L_r(A)$  = antisense strand likelihood of model A

$L_f(\text{max other})$  = the maximum likelihood over all models other than A for the sense strand

$L_r(\text{max other})$  = the maximum likelihood over all models other than A for the antisense strand

## Even Background Models

### Near Perfect Fit

A near perfect fit is when the best sense and antisense strand models occur for the same base. In this case, the quality scores for the sense strand,  $Q_f(\text{model})$ , and the antisense strand,  $Q_r(\text{model})$ , are both large (both strands fit the same model). If both strand quality scores for the model,  $Q(\text{model})$ , are greater than the quality threshold (default = 2), the model genotype is called for the site.

### Imperfect Fit

The imperfect fit is when the total quality score for a model is large, but one of the strands may not fit the model well. Sometimes the signal for one strand is weak, but is very strong for the other strand. An imperfect fit meets all of the following conditions:

$Q(\text{model}) > \text{Total quality threshold}$

$Q_f(\text{model}) \geq \text{Strand quality threshold}$

$Q_r(\text{model}) \geq \text{Strand quality threshold}$

The default total quality threshold is 75 and the default strand quality threshold is 0.

### No Call

If there is no model that fits the data near perfectly or imperfectly, the site is classified a no call (n).

## Uneven Background Models for Diploid Data

After analysis with the even background models, diploid data are further analyzed using refined genotype models (four homozygote, six heterozygote) that do not assume an even background mean and variance across all features. We derive an estimate for the background from the probe intensity variance across all samples in a model. The uneven background models assume that the background mean and variance are constant ratios of each other across all features. Two ratio constants,  $\alpha$  and  $\beta$ , are computed for each model and sample set. The analysis is an iterative process and runs the uneven background models until the calls converge.

### Calls

If a model meets the requirements for a nearly perfect or imperfect fit, a genotype is called for the site.

### No Call

If a model cannot be called or guessed, the site is classified a no call (n).

## Final Reliability Rules

After the algorithm makes the initial calls, three types of reliability rules are applied.

### Minimum Fraction of Calls in Neighboring Probes

In order for a call to be considered reliable, at least 50% of the surrounding sites (20 sites, 10 on each side) must be called. If there are >50% no calls in these 20 positions, the site is called n. The fraction is an adjustable threshold.

### Minimum Fraction of Calls in Samples

If >50% of the calls at particular position across all samples are no calls, the site is called n for all samples. The fraction is an adjustable threshold.

### Elimination of SNP Doublets

When two SNPs are within five bases of each other on a 25-mer probe, they are called a SNP doublet. A SNP doublet reduces probe reliability. Let SNP1 and SNP2 represent a SNP doublet; individuals homozygous for the reference base are called wild-type and all others are called mutant. The following rules determine the calls for SNP doublets.

1. If an individual is mutant at SNP1 and wild-type at SNP2, and another individual is wild-type SNP1 and mutant SNP2, both SNPs are reliable and the calls are unchanged.
2. If an individual is mutant at SNP1 and wildtype at SNP2, and all individuals mutant at SNP2 are mutant or n at SNP1, the SNP2 call is unreliable and all individuals are called n at this site. Similarly, if an individual is mutant at SNP2 and wildtype at SNP1, and all individuals mutant at SNP1 are mutant or n at SNP2, the SNP1 call is unreliable and all individuals are called n at this site
3. If mutants at SNP1 always occur in individuals who are also mutant or n at SNP2, and vice versa, the site with the smaller number of n calls is considered reliable. The other site is called n. If the two sites have an equal number of n calls, both sites are considered unreliable and are called n in all individuals.

## Heterozygosity

Heterozygosity is computed from the frequencies of the reference calls, A, and the mutant calls, B. The reference call frequency ( $p_A$ ) and mutant call frequency ( $p_B$ ) are calculated as:

$$p_A = (\text{Number of AA calls} + 0.5 * \text{Number of AB calls}) / \text{Total Number of Calls}$$

$$p_B = (\text{Number of BB calls} + 0.5 * \text{Number of AB calls}) / \text{Total Number of Calls}$$

where the Total Number of Calls does not include the no calls (n).

The heterozygosity (H) is calculated as:

$$H = 1 - (p_A^2 + p_B^2)$$

## SNP Frequency

The heterozygosity is calculated only from the frequencies of the mutant calls, B. The SNP frequency is calculated as:

$$P_B = (\text{Number of BB calls} + 0.5 * \text{Number of AB calls}) / \text{Total number calls}$$

## References

1. Cutler, D.J., Zwick, M.E., Carrasquillo, M.M., Yohn, C.T., Tobin, K.P., Kashuk, C., Mathews, D.J., Shah, N.A., Eichler, E.E., Warrington, J.A., and Chakravarti, A. 2001. High-Throughput Variation Detection and Genotyping Using Microarrays, *Genome Research*, 11:1913-1925.

## Resequencing Algorithm Version 1 Settings

You can modify settings for some of the Resequencing calling algorithm parameters.

There are several settings that affect the call rate and call accuracy of the algorithm. The default values are set for a reasonable call rate with high call accuracy. There is a trade off between the call rate and accuracy. As the call rate increases, the accuracy tends to decrease. You should exercise caution in changing these settings.

To view or change the user-modifiable algorithm settings:

1. In the Resequencing Analysis window, click the **AlgorithmParameters** button.

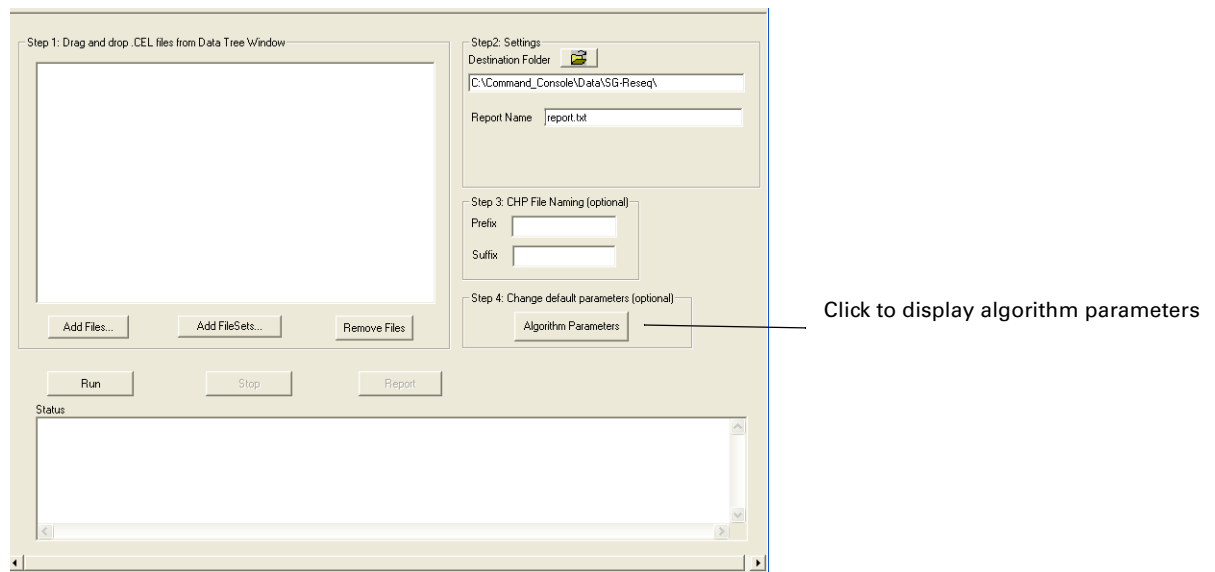


**Figure B.1** Resequencing Analysis window

The Resequencing Algorithm Settings dialog box opens (Figure B.2).

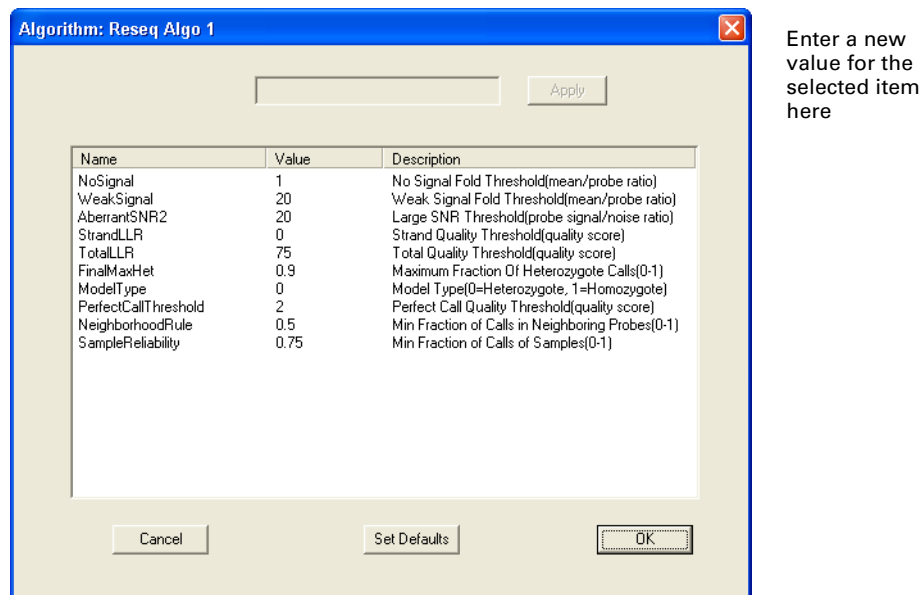


**Figure B.2** Resequencing Algorithm Version 1 Settings dialog box

2. To change a setting:
  - A. Click the item you want to change.  
The current value for the item is displayed.
  - B. Enter a new value for the item.
  - C. Click the **Apply** Button.
3. Click the **OK** button to set the new parameters.  
To return all settings to the defaults, click **Default**.

The settings are described in the following table.

**Table B.1** Resequencing Algorithm Version 1 Settings, user-modifiable data filters

| Name                 | Short Description                              | Longer Description                                                                                                                                                                                                                                                                                                                                                | Effect of Changing the Filter |                 |
|----------------------|------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|-----------------|
|                      |                                                |                                                                                                                                                                                                                                                                                                                                                                   | Filter Threshold Change       | Number of Calls |
| NoSignal             | No Signal Threshold (probe signal/noise ratio) | If any feature (in the sense or antisense strand) for a position has a mean intensity within two standard deviations of zero, the position is called n (no call).                                                                                                                                                                                                 | Increase                      | Decrease        |
| WeakSignal           | Weak Signal Fold Threshold (mean/probe ratio)  | If the highest mean intensity of a feature on the sense or antisense strand is 20-fold lower than the average highest mean intensity averaged over all samples on the same strand, the position is called n.                                                                                                                                                      | Increase                      | Increase        |
| AberrantSNR2         | Large SNR Threshold (probe signal/noise ratio) | SNR = Mean intensity of the feature/Standard deviation of the intensity for the feature. If the SNR for a position > 20, set the variance for the position so that SNR = 20.                                                                                                                                                                                      | N/A                           | N/A             |
|                      |                                                |                                                                                                                                                                                                                                                                                                                                                                   | Parameter Threshold Change    | Number of Calls |
| StrandLLR            | Strand Quality Threshold (quality score)       | See above.                                                                                                                                                                                                                                                                                                                                                        | Increase                      | Decrease        |
| TotalLLR             | Total Quality Threshold (quality score)        | Make a call if all of the following conditions are met:<br>1. Quality score of the sense strand > strand quality threshold.<br>2. Quality score of the antisense strand ≥ strand quality threshold.<br>3. Total quality score (sum of the sense and antisense strand quality scores) ≥ total quality threshold.<br>Otherwise, the position is called n (no call). | Increase                      | Decrease        |
| FinalMaxHet          | Maximum Fraction of Heterozygote Calls (0-1)   | After applying the uneven background models, if more than 90% of the calls are heterozygotes, set all features to n.                                                                                                                                                                                                                                              | Increase                      | Increase        |
| ModelType            | Model Type (0=heterozygote, 1=homozygote)      | Choose model type 0 for diploid data; choose model type 1 for haploid data.<br>Default = 0 for heterozygote.                                                                                                                                                                                                                                                      | N/A                           | N/A             |
| PerfectCallThreshold | Perfect Call Quality Threshold (quality score) | Make a call if all of the following conditions are met:<br>1. Quality score for the sense strand is greater than the Perfect Call Quality threshold.<br>2. Quality score for the antisense strand is greater than the Perfect Call Quality threshold.                                                                                                             | Increase                      | Decrease        |
|                      |                                                |                                                                                                                                                                                                                                                                                                                                                                   | Rule Change                   | Number of Calls |
| NeighborhoodRule     | Min Fraction of Calls in Neighboring Probes    | If <0.5 of the up to 20 features that surround a site (up to 10 features on each side) are not called, set the site to n (no call). Set the Min Fraction value to 1 to turn off the filter.                                                                                                                                                                       | Decrease fraction             | Decrease        |
| SampleReliability    | Min Fraction of Calls in Samples               | If a site is called n in >0.75 samples, set all samples to n at this site. Set the Min Fraction value to 1 to turn off the filter.                                                                                                                                                                                                                                | Decrease fraction             | Decrease        |

## RESEQUENCING ALGORITHM VERSION 2

The Resequencing algorithm, version 2, analyzes the cell intensity data (\*.cel) from an Affymetrix® resequencing probe array (8 µm) to provide the following results:

- Base calls with quality scores
- Heterozygosity
- SNP frequency

This appendix explains:

- How the Resequencing algorithm version 2 works (see [Resequencing Algorithm Version 2](#), below)
- How to adjust different parameters in the algorithm (see [Resequencing Algorithm Version 2 Settings on page 82](#))

### Resequencing Algorithm Version 2

This section contains information on the following topics:

- [Introduction](#)
- [Pre-Processing Filters on page 78](#)
- [Base Calling Algorithm on page 78](#)
- [Wild-type Sequence Profile and Trace on page 80](#)
- [Final Reliability Rules on page 81](#)
- [SNP Report Calculations on page 82](#)
- [References on page 82](#)

### Introduction

The Resequencing Algorithm, version 2, uses hypothesis testing over a set of models for haploid and diploid data with the following steps:

1. Pre-processing filters: checks for samples with weak or saturated cells
2. Calling algorithm: makes the base calls for each sample
3. Wild-type sequence profile and trace: detects signal reduction caused by mutant bases
4. Final reliability rules: evaluates wild-type sequence trace and apply sample reliability rule

The algorithm works best with multiple samples, and the user should choose a minimum of 15 independent samples (not technical replicates) for an analysis.

After the algorithm has determined the bases and SNPs in the samples, the heterozygosity and SNP frequencies are calculated for the SNP Reports (see [SNP Report Calculations on page 82](#)).

## Notation

An Affymetrix® resequencing probe array consists of a number of probe cells or features. During scanning, the software divides each probe cell into subunit squares or pixels. Each probe cell contains many copies of a unique 25-base oligonucleotide probe of defined sequence. Eight probe cells query a specific site in a known reference sequence. Four probe cells interrogate the sense strand and contain probes that are identical except for the central base which is A, C, G, or T. Four probe cells interrogate the antisense strand and contain probes that are identical except for the central base which is A, C, G, or T.

The Affymetrix GeneChip® Command Console (AGCC) computes the mean and standard deviation for a data file (\*.dat) and saves these to a cell intensity file (\*.cel).

## Pre-Processing Filters

The data filters identify unreliable data or adjust the extremely small variance of intensities that approach the detector limits.



**NOTE:** The filter thresholds and other parameters can be adjusted in the Reseq Algo 2 dialog box (see [Resequencing Algorithm Version 2 Settings on page 82](#)).

---

### No Signal

If any feature for a site (in the sense or antisense strand) has a mean intensity within two standard deviations of zero, the site is ruled a no call (n) for the sample.

### Weak Signal

If the highest mean intensity of a feature for the sense (or antisense) strand is 20-fold lower than the highest mean intensity feature, averaged across all samples, for the sense (or antisense) strand, the site is ruled a no call (n) for the sample.

### Saturation

If two probe cells (haploid data) or three probe cells (diploid data) in a four probe cell set for the sense or antisense strand are within two standard deviations of saturation level, the site is ruled a no call (n) for the sample.

### Maximum Signal-to-Noise Ratio

A signal with a very large signal-to-noise ratio (SNR) tends to have an exaggerated influence on the calling algorithm. If the SNR is greater than 20, the variance is set so that the SNR = 20.

## Base Calling Algorithm

The calling algorithm uses the following heuristic to make the base calls

1. For every base in each sample, the maximum likelihood estimate is calculated for each of the different models.
2. The likelihoods for each model are compared and a quality score is calculated to make a preliminary call for each sample.
3. Iterative adaptive background adjustments are performed to compensate for non-uniform background cell intensities.
4. The final base call is made using the base calls and quality scores for the forward and reverse strands.

## Calculating Log Maximum Likelihood Estimates

The Resequencing calling algorithm is based on the Adaptive Background Genotype Calling Scheme (ABACUS) developed by Cutler et al. (1). This algorithm specifies models for the presence or absence of various genotypes in the sample. For haploid data there are five genotype models (A, C, G, T, no call) and for diploid data there are 11 genotype models (A, C, G, T, AC, AG, AT, CG, CT, GT, no call).

The models assume the pixel intensities of a feature are independently and normally distributed. The algorithm computes the estimated mean background and variance for the sense and antisense strand features. Initially, the data are analyzed using even background models that assume the same mean background intensity and variance across all features.

Each model generates an estimate of feature intensity and variation based on a likelihood model. The likelihood of each model is computed as:

$$ll = -\frac{1}{2} \sum_x N_x [\ln(\hat{\sigma}_x^2) + (V_x + M_x^2 - 2\hat{\mu}_x M_x + \hat{\mu}_x^2) / \hat{\sigma}_x^2 + \ln(2\pi)]$$

where:

$N_x$  = the number of pixels observed in feature x

$V_x$  = the observed variance for feature x

$M_x$  = observed mean for feature x

$\hat{\mu}_x$  = estimated mean for feature x under the model in question

$\hat{\sigma}_x^2$  = estimated variance for feature x

The sum is taken over all features x, where x is A, C, G, or T, on the sense and antisense strand features. Note that  $\hat{\mu}_x$  and  $\hat{\sigma}_x^2$  differ for each model.

The algorithm computes the likelihood of a particular genotype model independently for the sense strand and antisense strand.

## Quality Score

The log maximum likelihood scores for each base and model are used to calculate the quality score:

$$RS(M) = ll(M) - \max\{ll(m), m \in S, m \neq M\}$$

Where M is the best fitting model for the strand.

Two strand quality scores,  $RSf$  and  $RSr$ , are calculated independent for the forward and reverse strands.

Notice that the best fitting model for one strand may not be the best fitting model for the other, and this is also true for the second best fitting model.

## Adaptive Background Adjustments

The background cells are often not uniform. Cross hybridization, chip variation, and other factors make the background vary from base to base and from sample to sample. These variations can be quite large and make the even background a poor assumption. The adaptive background procedure is designed to reflect the background differences between probes and samples. The adaptive background adjustment has a set of coefficients associated with different background bases.

After analysis with the even background models, diploid data are further analyzed using refined genotype models (four homozygote, six heterozygote) that do not assume an even background mean and variance across all features. We derive an estimate for the background from the probe intensity variance across all samples in a model. The uneven background models assume that the background mean and variance are constant ratios of each other across all features. Two ratio constants,  $\alpha$  and  $\beta$ , are computed for each model and sample set. The analysis is an iterative process and runs the uneven background models until the calls converge.

## Final Base Calling (combine strand)

The final base calling uses the base calls and quality scores for the forward and reverse strands according to the following set of rules, applied in order:

1. If both forward and reverse strands are filtered out by the preprocessing filters, no call is assigned to this base and quality score is zero.
2. If one strand is filtered out, use the other strand for both base call and quality score.
3. If the base call is the same on both strands, use the common call as base call and assign quality score as the sum of the two strand's quality score.
4. If the two strands are different heterozygote calls, no call is made and quality score is zero.
5. If one strand's call is homozygote, and the other strand's call is heterozygote, and the homozygote call is one of the two in the heterozygote call, use the strand with the homozygote call for both call and quality score
6. Else, no call is made and zero is assigned as the quality score.
7. A quality score threshold is set to filter out calls with low quality scores, if  $RS < \text{quality score threshold}$  change call to no call.

For more information about changing this parameter see [Resequencing Algorithm Version 2 Settings](#) on page 82.

## Wild-type Sequence Profile and Trace

The probe sequence is optimized to the wild type reference sequence. The mutation call has one base that is different from the reference sequence. A mutation call tends to reduce the intensity near neighboring bases.

This can result in a lower probe set affinity and cause a reduction of signal with the neighboring probes ([Figure C.1](#)).

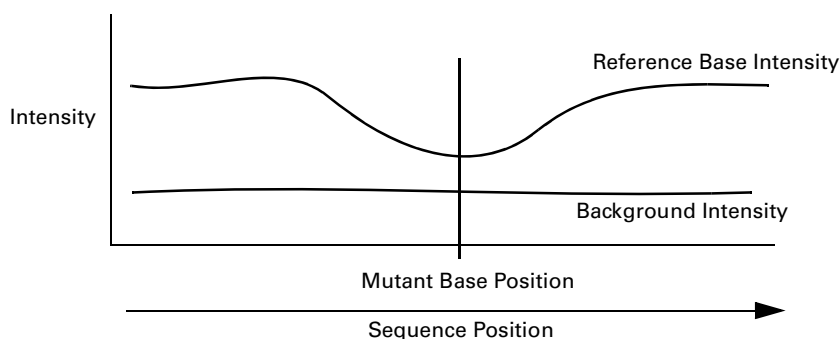


**Figure C.1** Intensities with a mutant base

The trace profile metrics are designed to detect this signal reduction.

To define the wild-type sequence profile and trace:

1. Take a Center Exponentially Weighted Moving Average (CEWMA) of the feature signals with reference bases to smooth the wild-type signals; the average is taken over a limited size window moving along the sequence.
2. Smooth the signals with “potential” mutant bases; a “potential” mutant base is defined as the feature with the highest signal among the three features other than the reference base
3. The wild-type sequence profile is defined as the log ratio of the wild-type CEWMA and the “potential” mutant CEWMA.
4. Calculate the trace as the second derivative of the wild-type sequence profile.

### Calculate the Wild-type and “Potential” Mutant Average

Let  $x$  be the index of a given base, the wild-type smoothed signal and “potential” mutant smoothed signal are defined as:

$$I_{wt}^s(x) = \sum_{i=-k}^k \alpha(1-\alpha)^{|i|} I_{wt}^o(x+i)$$

$$I_{mt}^s(x) = \sum_{i=-k}^k \alpha(1-\alpha)^{|i|} I_{mt}^o(x+i)$$

Where:

s = smoothed

o = observed

wt = wild-type

mt = mutant

I = signal intensity

$I_{mt}^o(y) = \max\{I_k^o(y) \mid k \in \{A, C, G, T\} \text{ and } k \text{ is not the reference base}\}$

and  $\alpha$  is a constant, default value is 0.5

### Calculate the Wild-Type Sequence Profile

The wild-type sequence profile  $F(x)$  is defined as:

$$F(x) = \log\left(\frac{I_{wt}^s(x)}{I_{mt}^s(x)}\right)$$

### Calculate the Wild-Type Trace

The wild-type trace uses the second derivative of the wild type sequence profile. The wild-type trace is defined as:

$$Tr(x) = \frac{1}{2}[-F(x-\Delta x) + 2F(x) - F(x+\Delta x)]$$

Where  $\Delta x$  is the step size for the derivative and default value is 2.

The wild-type sequence profile and trace values are used in the final quality check (see below).

## Final Reliability Rules

A set of final reliability rules are applied to confirm that either the base call or a genotype call (mutation) is detected by the calling algorithm.

### Trace rule

For any genotype call at base  $x$ , if:

$F(x) > \text{Sequence Probe Threshold}$  and  $Tr(x) > \text{Trace Threshold}$ ,

then overwrite the original call by no call.

For more information about changing these settings see [Resequencing Algorithm Version 2 Settings on page 82](#).

### Sample Reliability Rules

For any base, if the base call rate is under the threshold across all samples, overwrite all calls on this base by no call

For more information about changing this setting see [Resequencing Algorithm Version 2 Settings on page 82](#).

## SNP Report Calculations

The software also calculates:

- Heterozygosity
- SNP Frequency

These values are reported in the SNP Report (see [Contents of the Resequencing SNP Report on page 72](#)).

### Heterozygosity Calculation

Heterozygosity is computed from the frequencies of the reference calls, A, and the mutant calls, B. The reference call frequency ( $p_A$ ) and mutant call frequency ( $p_B$ ) are calculated as:

$$p_A = (\text{Number of AA calls} + 0.5 * \text{Number of AB calls}) / \text{Total Number of Calls}$$

$$p_B = (\text{Number of BB calls} + 0.5 * \text{Number of AB calls}) / \text{Total Number of Calls}$$

where the Total Number of Calls does not include the no calls (n).

The heterozygosity (H) is calculated as:

$$H = 1 - (p_A^2 + p_B^2)$$

### SNP Frequency

The heterozygosity is calculated only from the frequencies of the mutant calls, B. The SNP frequency is calculated as:

$$P_B = (\text{Number of BB calls} + 0.5 * \text{Number of AB calls}) / \text{Total number calls}$$

## References

1. Cutler, D.J., Zwick, M.E., Carrasquillo, M.M., Yohn, C.T., Tobin, K.P, Kashuk, C., Mathews, D.J., Shah, N.A., Eichler, E.E., Warrington, J.A., and Chakravarti, A. 2001. High-Throughput Variation Detection and Genotyping Using Microarrays, *Genome Research*, 11:1913-1925.

## Resequencing Algorithm Version 2 Settings

You can modify settings for some of the algorithm parameters for Resequencing algorithm, version 2.

There are several settings that affect the call rate and call accuracy of the algorithm. The default values are set for a reasonable call rate with high call accuracy. There is a trade off between the call rate and accuracy. As the call rate increases, the accuracy tends to decrease. You should exercise caution in changing these settings.

**To view or change the user-modifiable algorithm settings:**

1. In the Resequencing Analysis window, click the **AlgorithmParameters** button.

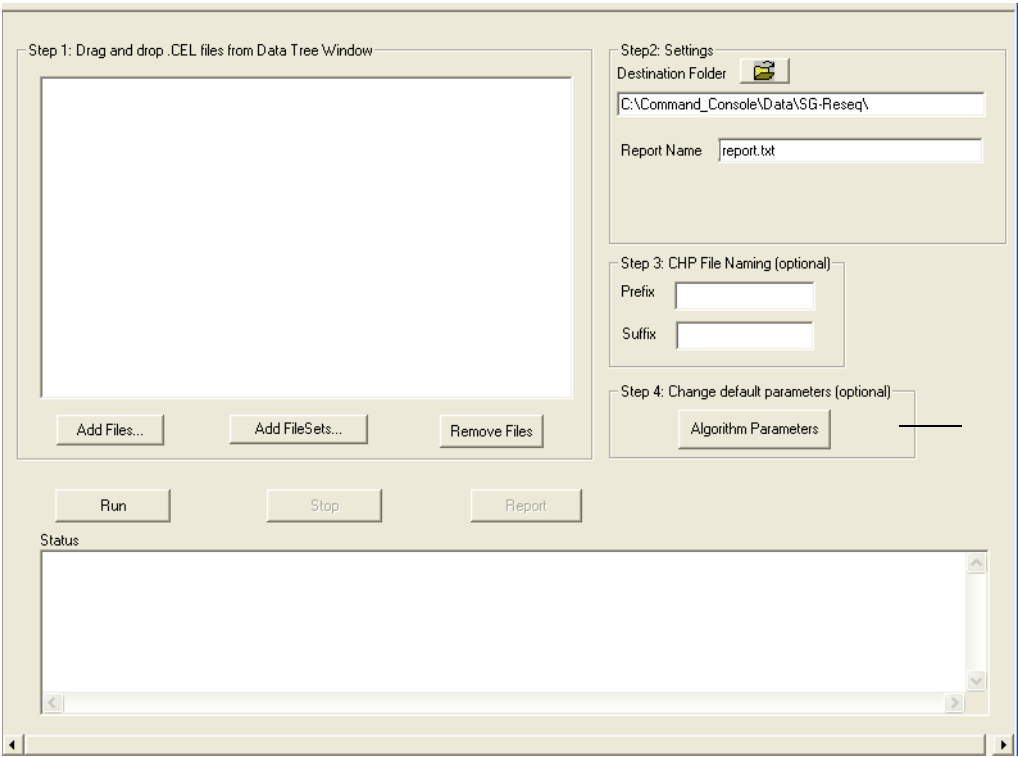


Figure C.2 Resequencing Analysis window

The Resequencing Algorithm Version 2 Settings dialog box opens (Figure C.3).

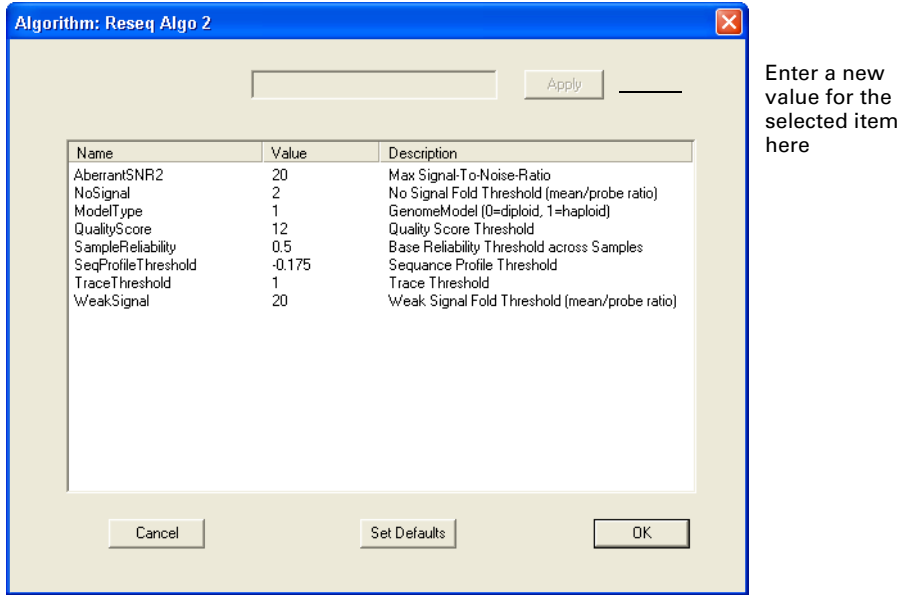


Figure C.3 Resequencing Algorithm Version 2 Settings dialog box

- 2. To change a setting:
  - A. Click the item you want to change.  
The current value for the item is displayed (Figure C.3).
  - B. Enter a new value for the item.

**C.** Click the **Apply** Button.

**3.** Click the **OK** button to set the new parameters.

To return all settings to the defaults, click **Default**.

The settings are described in the following table.

**Table C.1** Resequencing Algorithm Version 2 Settings, user-modifiable data filters

| Name                 | Short Description                             | Description                                                                                                                                                                                                                  | Effect of Changing the Value |                 |
|----------------------|-----------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|-----------------|
|                      |                                               |                                                                                                                                                                                                                              | Filter Threshold Change      | Number of Calls |
| Aberrant SNR2        | Max Signal-To-Noise-Ratio                     | SNR = Mean intensity of the feature/Standard deviation of the intensity for the feature. If the SNR for a position > 20, set the variance for the position so that SNR = 20.                                                 | N/A                          | N/A             |
| NoSignal             | No Signal Fold Threshold*                     | If any feature (in the sense or antisense strand) for a position has a mean intensity within two standard deviations of zero, the position is called n (no call).                                                            | Increase                     | Indeterminate   |
| ModelType            | Genome Model (0=diploid, 1=haploid)           | Choose type 0 for diploid data; choose type 1 for haploid data. Default = 0 for heterozygote.                                                                                                                                | <b>N/A</b>                   | <b>N/A</b>      |
| QualityScore         | Quality Score Threshold                       | Make a call if quality score is greater than the threshold.                                                                                                                                                                  | <b>Increase</b>              | <b>Decrease</b> |
| Sample Reliability   | Base Reliability Threshold across Samples     | If the percentage of samples with a no call for a particular base position is less than the threshold (expressed as a fraction), all samples are set to N at that position. Setting the threshold to 0 turns off the filter. | <b>Increase</b>              | <b>Decrease</b> |
| SeqProfile Threshold | Sequence Profile Threshold                    | Overwrite original call by no call if Sequence profile is greater than the threshold.                                                                                                                                        | <b>Increase</b>              | <b>Increase</b> |
| Trace Threshold      | Trace Threshold                               | Overwrite original call by no call if trace is greater than threshold.                                                                                                                                                       | <b>Increase</b>              | <b>Increase</b> |
| Weak Signal*         | Weak Signal Fold Threshold (mean/probe ratio) | If the highest mean intensity of a feature on the sense or antisense strand is 20-fold lower than the average highest mean intensity averaged over all samples on the same strand, the position is called n (no call).       | Increase                     | Indeterminate   |

N/A = Not Applicable

\*Lowering the No Signal or Weak Signal parameter can either increase or lower the call rate, depending upon which of two competing effects dominate when lowering the threshold. The effects are:

- 1) Increases the number of bases with no signal or weak signal calls on both strands. This lowers the call rate.
- 2) Increases the number of bases with no signal or weak signal calls on one strand. In this case, the call is only made on the second strand. This can increase the call rate.

FILE TYPES

Probe Information Files

The probe information or library files contain information about the probe array design characteristics, probe utilization and content, and scanning and analysis parameters. These files are unique for each probe array type.

Sample and Data Files

The following file types are used or generated by GSEQ (see [Table D.1](#)).

Table D.1 Affymetrix file types

| Experiment Data<br>File Name                    | Created in... | Description                                                                                                                                                                                                                                                                                                                                                                                                              |
|-------------------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sample (.ARR) file)                             | AGCC          | The beginning of the data chain for a given experiment. The sample information is stored in a Sample file with an ARR extension. The arrays used in analysis and data files produced by analysis are linked to this Sample File.<br>The information about the sample and experiment are collected as attributes. These attributes can then be used to locate particular Sample files in filtering and search operations. |
| Data file (*.DAT)                               | AGCC          | The image of the scanned probe array.                                                                                                                                                                                                                                                                                                                                                                                    |
| Cell intensity data<br>file (*.CEL)             | AGCC          | The software derives the *.cel file from a *.dat file and automatically creates it upon opening a *.dat file. The *.cel contains a single intensity value for each probe cell delineated by the grid (computed by the Cell Analysis algorithm).                                                                                                                                                                          |
| Analysis results or<br><i>chip</i> file (*.CHP) | GSEQ          | The analysis results file.                                                                                                                                                                                                                                                                                                                                                                                               |
| Report file (*.RPT)                             | GSEQ          | Report files used to provide additional information about analysis algorithm and results.                                                                                                                                                                                                                                                                                                                                |
| Data file<br>(*.TXT)                            | GSEQ          | *.txt is a standard format for text files. GSEQ exports and imports this file format.                                                                                                                                                                                                                                                                                                                                    |



IUPAC BASE CODES

Table E.1 IUPAC Base Codes

| IUPAC Code | Group           | Base(s)                            |
|------------|-----------------|------------------------------------|
| A          | A               | Adenine                            |
| C          | C               | Cytosine                           |
| G          | G               | Guanine                            |
| T          | T               | Thymine                            |
| M          | A or C          | aMino                              |
| R          | A or G          | puRine                             |
| W          | A or T (U)      | Weak interaction (2 H bonds)       |
| Y          | C or T (U)      | pYrimidine                         |
| S          | C or G          | Strong interaction (3 H bonds)     |
| K          | G or T(U)       | Keto                               |
| V          | A or C or G     | not-T or not-U (since V follows U) |
| H          | A or C or T(U)  | not-G (since H follows G)          |
| D          | A or G or T(U)  | not-C (since D follows C)          |
| B          | C or G or T(U)  | not-A (since B follows A)          |
| N          | A, C, G or T(U) | aNy                                |



## WORKING WITH WINDOWPANES & COLUMNS

### Resizing Windowpanes

You can resize windowpanes.

- Place the mouse pointer over a windowpane border so that it changes to a double arrow  $\leftrightarrow$  (Figure F.1).

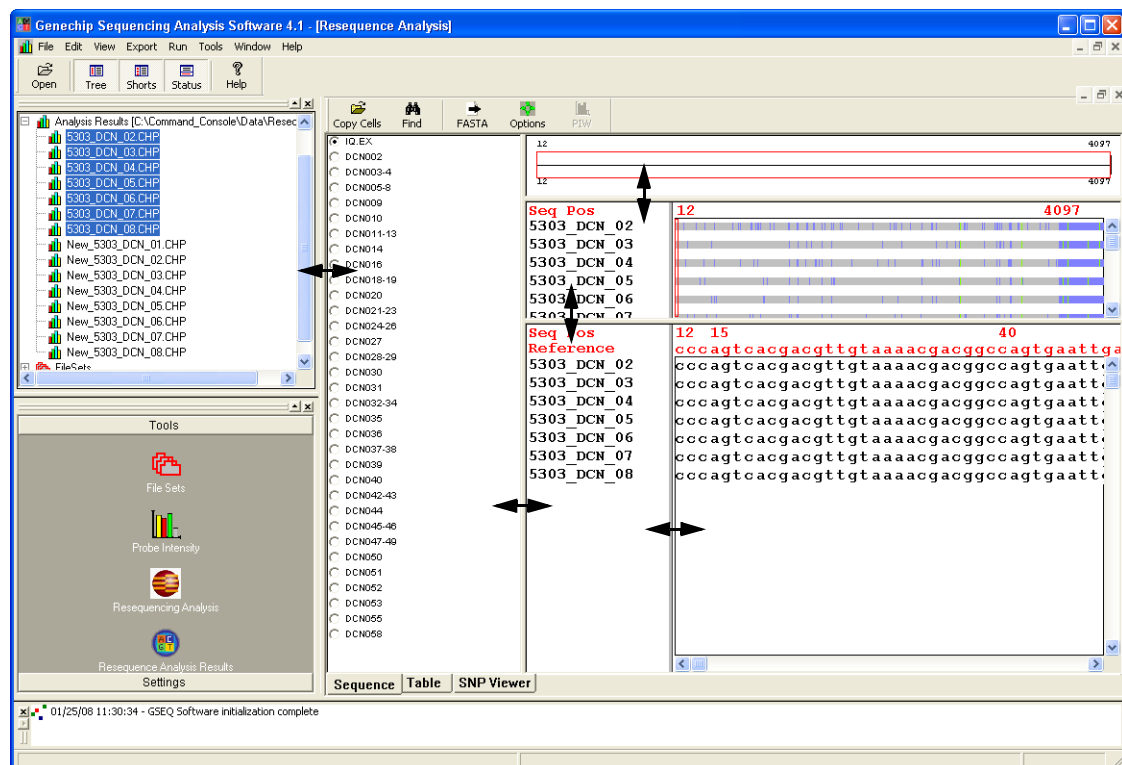


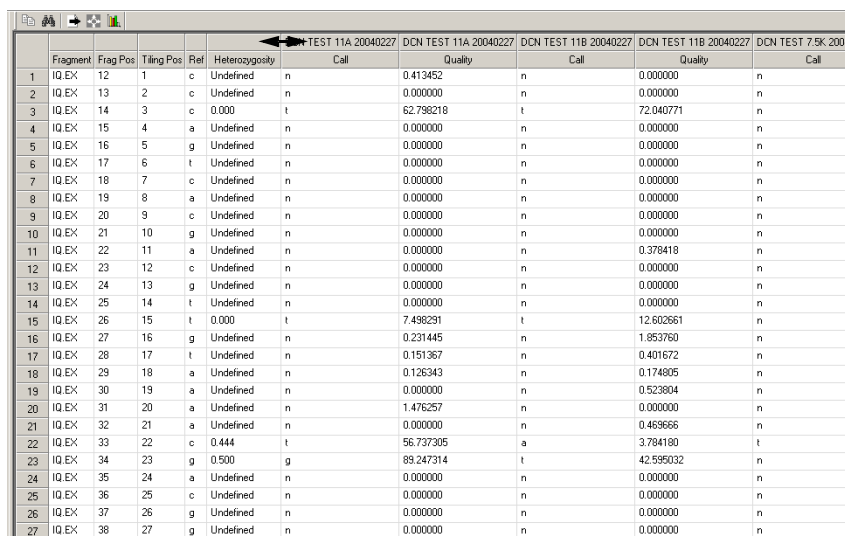
**Figure F.1** Resize windowpanes horizontally or vertically

1. Drag the border to resize the windowpane.

## Resizing or Hiding Table Columns

You can resize or hide columns in the table view.

1. Position the mouse over the left or right cell border in the column header so that it changes to a double arrow  $\longleftrightarrow$ .
2. Drag the cell border to resize the width of the column.
3. To hide a column, drag the left or right cell border of the column header until the column width is reduced to zero, or right-click the column header in the table and click **Hide Column** in the shortcut menu.



|    | Fragment | Frag Pos | Tiling Pos | Ref | Heterozygosity | DCN TEST 11A 20040227 | DCN TEST 11A 20040227 | DCN TEST 11B 20040227 | DCN TEST 11B 20040227 | DCN TEST 7.9K 2004 |
|----|----------|----------|------------|-----|----------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------|
|    |          |          |            |     |                | Call                  | Quality               | Call                  | Quality               | Call               |
| 1  | IQ.EX    | 12       | 1          | c   | Undefined      | n                     | 0.413452              | n                     | 0.000000              | n                  |
| 2  | IQ.EX    | 13       | 2          | c   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 3  | IQ.EX    | 14       | 3          | c   | 0.000          | t                     | 62.796218             | t                     | 72.040771             | n                  |
| 4  | IQ.EX    | 15       | 4          | a   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 5  | IQ.EX    | 16       | 5          | g   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 6  | IQ.EX    | 17       | 6          | t   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 7  | IQ.EX    | 18       | 7          | c   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 8  | IQ.EX    | 19       | 8          | a   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 9  | IQ.EX    | 20       | 9          | c   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 10 | IQ.EX    | 21       | 10         | g   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 11 | IQ.EX    | 22       | 11         | a   | Undefined      | n                     | 0.000000              | n                     | 0.378418              | n                  |
| 12 | IQ.EX    | 23       | 12         | c   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 13 | IQ.EX    | 24       | 13         | g   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 14 | IQ.EX    | 25       | 14         | t   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 15 | IQ.EX    | 26       | 15         | t   | 0.000          | t                     | 7.496291              | t                     | 12.602661             | n                  |
| 16 | IQ.EX    | 27       | 16         | g   | Undefined      | n                     | 0.231445              | n                     | 1.853760              | n                  |
| 17 | IQ.EX    | 28       | 17         | t   | Undefined      | n                     | 0.151367              | n                     | 0.401672              | n                  |
| 18 | IQ.EX    | 29       | 18         | a   | Undefined      | n                     | 0.126343              | n                     | 0.174805              | n                  |
| 19 | IQ.EX    | 30       | 19         | a   | Undefined      | n                     | 0.000000              | n                     | 0.523804              | n                  |
| 20 | IQ.EX    | 31       | 20         | a   | Undefined      | n                     | 1.476257              | n                     | 0.000000              | n                  |
| 21 | IQ.EX    | 32       | 21         | a   | Undefined      | n                     | 0.000000              | n                     | 0.469666              | n                  |
| 22 | IQ.EX    | 33       | 22         | c   | 0.444          | t                     | 56.737305             | a                     | 3.784180              | t                  |
| 23 | IQ.EX    | 34       | 23         | g   | 0.500          | g                     | 89.247314             | t                     | 42.595032             | n                  |
| 24 | IQ.EX    | 35       | 24         | a   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 25 | IQ.EX    | 36       | 25         | c   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 26 | IQ.EX    | 37       | 26         | g   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |
| 27 | IQ.EX    | 38       | 27         | g   | Undefined      | n                     | 0.000000              | n                     | 0.000000              | n                  |

Figure F.2 Resequencing Analysis window, table view

**HOT KEY DESCRIPTIONS**

**Table G.1** Hot key descriptions

| Menu Bar Command                    | Hot Key  |
|-------------------------------------|----------|
| File → Open                         | Ctrl + O |
| File → Print                        | Ctrl + P |
| Edit → Cut                          | Ctrl + X |
| Edit → Copy Cells or<br>Edit → Copy | Ctrl + C |
| Edit → Paste                        | Ctrl + V |
| Edit → Image Settings               | S        |
| Edit → Delete                       | Del      |
| Edit → Find Next                    | F3       |
| Help                                | F1       |



# INDEX

## Symbols

- .cel extension 85
  - See also cell intensity data
- .CEL files
  - see CEL files
- .chp extension 85
  - See also analysis results file (\*.chp)
- .CHP files
  - see CHP files
- .dat extension 85
- .DAT files
  - see DAT files
- .exp extension 85
- .rpt extension 85
  - See also reports

## A

- Add Cell/Chip data items dialog box 23
- Affymetrix technical support 4
- algorithm 21
  - automatically selected for assay type 21
  - Resequencing 71, 77
- algorithm report 21
  - resequencing 61
  - See also Report window
- analysis results file (\*.chp) 85
  - See also analysis results
  - view file information 14
- analyzing cell intensity data 21
- Assay type
  - See also
  - Resequencing Arrays

## B

- base calls 45
- base codes
  - See IUPAC base codes

## C

- cell intensity data
  - analyzing
    - See analyzing cell intensity data
  - file type (\*.cel) 85
  - view file information 14
- CHP files 8

## D

- DAT files 8
- data export

- for Probe Intensity window 59
- for Resequencing analysis
  - FASTA format 35
  - tabular format 35

## data file

- \*.dat 85
- \*.txt 85

## data files 8

- see also
  - CEL files
  - CHP files
  - DAT files
  - parent-child file relationships

## data filters

- in Resequencing algorithm 71, 76, 78, 84

## data tree 13

- controlling display of 13

## documentation 4

- conventions used 3
- online 4

## E

- editing sequence information 37
- experiment information file (\*.exp) 85
- Export As dialog box 35
- exporting data
  - for Probe Intensity window 59
  - for Resequencing analysis
    - FASTA format 35
    - tabular format 35

## F

## FASTA Export dialog box 35

## feature 71, 78

## file types 8, 85

- cell intensity data (\*.cel) 85
- chip (\*.chp) 85
- data (\*.dat) 85
- data (\*.txt) 85
- experiment information (\*.exp) 85
- probe information (library) 85
- report file (\*.rpt) 85

## files

- data 8
- opening 17
- see also
  - CEL files
  - CHP files
  - DAT files
- viewing information 14

## FileSets 14

- creating 15

- editing 16
- renaming 16

## G

- GSEQ 1
  - installation 67
  - starting 8
- GSEQ help 4
- GTYPE 1
  - starting 8
- GTYPE help 4

## H

- help 4
- heterozygosity
  - Resequencing calling algorithm
  - heterozygosity 74, 82

## I

- Image (DAT) files
  - see DAT files
- installing GSEQ 67
- Intensity (CEL) data files
  - see CEL files
- intensity data
  - in Probe Intensity window 53
- Intensity Plot 55
- IUPAC base codes 87
- IUPAC warning box 38

## M

- main display area 12

## O

- online documentation 4
- on-line help 4
- opening files 17

## P

- printing
  - sequence maps 46
- Probe Analysis (CHP) files
  - see CHP files
- probe cell 71, 78
  - in Probe Intensity Window 53
- probe information files 85
- Probe Intensity table 58
- Probe Intensity window 53
  - exporting data 59
  - Intensity Plot 55
  - Probe Intensity table 58
  - toolbar 55
- probe set intensity data
  - viewing
    - in Probe Intensity window 53

## R

- report file (\*.rpt) 85
- Report window 61
- reports 61
  - Resequencing Algorithm report 61
- resequence analysis
  - See Resequencing analysis
- Resequencing Analysis window 27
  - Sequence view 40
  - Table View
    - editing sequence information 37
  - Table view 30
  - toolbar 29
- Resequencing algorithm 71, 77
  - data filters 71, 78
  - DNA analysis models 72
  - reliability rules 73, 81
  - settings 74, 82
    - data filters 76, 84
  - SNP frequency 74, 82
- Resequencing analysis 21
  - See also analyzing cell intensity data
- Resequencing Analysis Options dialog box 46
- Resequencing Arrays 2, 21
- resequencing results
  - base calls 45
  - editing sequence information 37
- resizing windowpanes 89

## S

- Save Chip dialog box 39
- screen captures 3
- sequence maps 44
  - printing 46
- Sequence view 40
- sequence view
  - options 46
  - sequence maps 44
  - windowpanes 40
- shortcut bar 17
  - Settings 18
  - Tools 18
- SNP frequency 74, 82
- Sort dialog box 32
- starting GSEQ 8
- starting GTYPE 8
- status log 18

## T

- Table view 30
  - copying 36
  - export to FASTA format 35
  - export to tabular format 35
  - hiding columns 33, 90
  - resizing 90
  - sorting 32–33
- technical support 4
- toolbar 12

- main 12
- Probe Intensity window 55
- Resequencing Analysis window 29

Tools shortcut bar 18

## **U**

- user interface 9, 11
  - components 9, 11
  - data tree 13
  - main display area 12
  - overview 11

## **W**

- window
  - Probe Intensity 53
  - Report 61
- windowpanes, resizing 89

